UNDERSTANDING NEURAL MECHANISMS OF STRATEGIC LEARNING: CORRELATES, CAUSALITY, AND APPLICATIONS

BY

LUSHA ZHU

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Economics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Assistant Professor Ming Hsu, Chair
Professor Steven R. Williams
Professor Nicholas Yannelis
Associate Professor Thomas J. Anastasio

ABSTRACT

This is a systematic study on learning in the repeated game from the neuroeconomics perspective. Theoretically, learning theory has been developed to complement the traditional game theory in seeking to explain how and which equilibria might arise as a consequence of nonequilibrium dynamics among agents with bounded rationality. Empirically, learning models have been widely used to describe the evolvement of observed behavior over the course of field and laboratory experiments. While game theorists are trying to make learning theory more empirically relevant (Fudenberg and K. Levine 2009), experimentalists often found it difficult to distinguish different learning models based on behavioral choice data alone (Salmon 2001; Wilcox 2006). Here I sought to investigate learning mechanism from an alternative perspective: the neuroeconomics perspective, by combining the game theory experimental paradigm, parametric learning models, and neuroscience methods.

In the first part of the thesis, I sought to identify the underlying learning rule by investigating how the brain encodes and computes learning signals used to guide behavior in a repeated normal-form game. Specifically, I combined functional neuroimaging of a multi-strategy competitive game with computational modeling of three widely used classes of learning models—reinforcement, belief-based learning, and their hybrid, experience-weighted attraction (EWA). I found evidence for distinct signals for reinforcement and belief-based learning in the brain. More importantly, I rejected the hypothesis of a hybrid EWA process at the neural level, even though it outperforms reinforcement and belief-based learning models behaviorally. Based on these findings, I hypothesized that behavioral choices are a product of a dual-system process at the brain level involving reinforcement and belief-based learning signals.

Although the neural imaging method provides a new dimension of data and biologically plausible criterion for model testing, it is silent about the causal relation between brain regions and learning signals. In order to validate the neuroimaging results and establish the necessary roles of brain regions for strategic learning, I then compared

the behavior of focal brain lesion patients to normal volunteers that are matched in terms of demographics and cognitive measures. In particular, I studied three different types of lesion patients: orbital frontal, dorsal lateral prefrontal and basal ganglia patients, which allowed me to dissociate the different roles necessarily to strategic learning.

In the third part of the thesis, I applied the above findings on the neural circuitry underlying strategic learning to explore the behavioral signature of a special yet important population, the elderly individuals. In particular, I compared the behavioral results from the strategic learning under two experimental settings: playing against other intelligent players and against a computer agent; and between two populations: the healthy elderly individuals and young individuals. Our behavioral results suggest that elderly individuals adjust more slowly. Interestingly, this is not because elderly individuals are insensitive to the new experience but because their prior belief decays more slowly than young individuals. I further posited that within elderly population, their prior decays more slowly when they are playing against intelligent people than against a computer agent. This comparative study serves as a first step for developing biomarkers to quantify decision-making deficits and will shed light on the individual differences in productivity and intellectual viability often found within the elderly population.

ACKNOWLEDGMENTS

My deepest gratitude goes to my advisor, Prof. Ming Hsu, who has introduced me into this very new field of neuroeconomics, and guided me through the periods of skepticism, confusion and struggles. He is a great advisor, supportive, accessible, and with enormous enthusiasm and intelligence on science. I am indebted to him for his insightful direction of the projects and the stimulating discussions.

I am grateful to Professor Steve Williams, Thomas J. Anastasio, Nicholas Yannelis, Bob Knight and Rick Ivy for the insightful comments and constructive criticism. I would also thank Professor In-Koo Cho and Rui Zhao whom I have been working with before I was introduced to neuroeconomics. Many thanks to my awesome collaborators, Eric Set, Mahdi Rastad, Dan Walsh and Kyle Mathewson, for all the cooperation, inspiration, and discussion at different stages of my research. Thanks to Nancy Dodge and Holly Tracy at UIUC, Donatella Scabini, Jessica Black and Clay Clayworth at Berkeley for the excellent services that made the research projects possible. Sincerest thanks to all my friends who supported me emotionally along the way.

Last but not least, I want to thank my family who endured this long process with me. Thanks to my parents for their love and unconditional supports though we are separated by thousands of miles; my cat, Ququer, for all the days and nights she spent besides me purring while I was working on the dissertation; and finally my husband, Feng Liu, for all the lousy jokes he collected from the Internet to cheer me up when I needed a break, and for the time we spent together and the time we spent missing each other.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1. Learning

Learning in Economics

Notions of equilibrium are central to theories of strategic behavior in the biological and social sciences where individuals must take into account not only available rewards and punishments but also behavior of other intelligent agents. It has long been recognized, however, that equilibria do not emerge spontaneously, but rather through some adaptive process whereby organisms evolve or learn over time.

In the field of economics, learning theory has been developed theoretically in complement to the traditional game theory seeking to explain how and which equilibria might arise as a consequence of nonequilibrium dynamics among players with bounded rationality. In particular, it develops adaptive models that are simple in implementation with some behavioral foundations, and investigates whether such unsophisticated, bounded rational learning rules can converge to behavior that is highly sophisticated and rational asymptotically. Furthermore, it also seeks to solve some of the long-standing conceptual problems in equilibrium theory, such as the absence of explanation for how a particular equilibrium is reached in the presence of multiple equilibria, and serves as an important way of evaluation the traditional equilibrium concepts (Fudenberg and K. Levine 1998; Fudenberg and K. Levine 2009; Foster and Vohra 1998; Hart and Mas-Colell 2003; Foster and Young 2006).

Empirically, learning models have been widely used to describe the evolvement of observed behavior in real economic lives as well as in field and lab experiments. For

example, macroeconomists explained the raise and fall of US inflation rates over the past 50 years by incorporating the adaptive learning process about the true Phillips curve into traditional econometric method of policy evaluations to explain Fed's dynamic choices over target inflation rates (Sargent 2006). In the field of microeconomics and industrial organization, it has also been shown that many types of consumer purchasing behavior can be better explained by taking into account the dynamic learning processes itself (Crawford and Shum 2005; Narayanan and Manchanda 2009).

Empirical Difficulty

A large number of theoretical models of learning have therefore been proposed. Consequently, an empirical question has come to the fore: which learning models describe human behavior best? Over the past decade, there has been a growing literature focusing on running statistical horse races among different learning models using observed choice data in order to identify which among these learning rules appear to be used (Roth and Erev 1995; Mookherjee and Sopher 1997; Cheung and Friedman 1998; Erev and Roth 1998; Erev, Roth et al. 2007). Laboratory studies, by employing tight control over the decision environment of subjects, provide perhaps the best-case scenario for identifying learning processes underlying choice behavior. Even in these ideal circumstances, however, researchers have had great difficulty to accurately distinguish between the actual underlying learning rules based on stimulated or observed choice data. Experimental economists showed that individual heterogeneity in learning would lead to severely biased estimations (Wilcox 2006), as illustrated in Figure 2. Moreover, Salmon fitted various learning models with the stimulated choice data and found that a number of

learning rules besides the true data generating process can provide statistically satisfactory fit, resulting in large Type I or Type II error when testing for the true model.

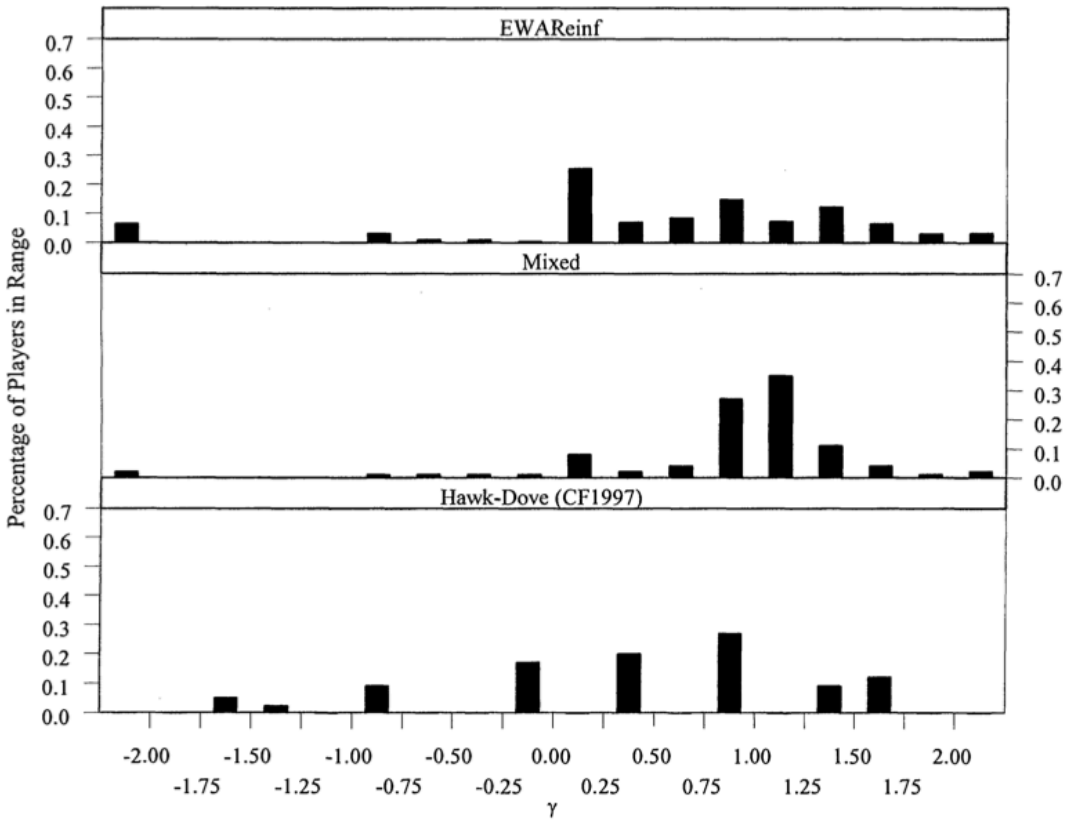

Figure 1: Histograms showing the frequency distribution of the estimated value for a key parameter for belief-based learning model in (Cheung and Friedman 1997), based on simulated reinforcement learning and mixed population players as well as a sample from Cheung and Friedman's original paper. In reinforcement learning the value should take value 0 whereas in belief based learning it should be 1. Hence ideally we will expect the estimated value for this parameter is near 0 for simulated reinforcement players and 1 for belief-based learners. But as shown in this figure, the chance for misclassifying the reinforcement learner as belief-based learner is very high. Adapted from Salmon (2001).

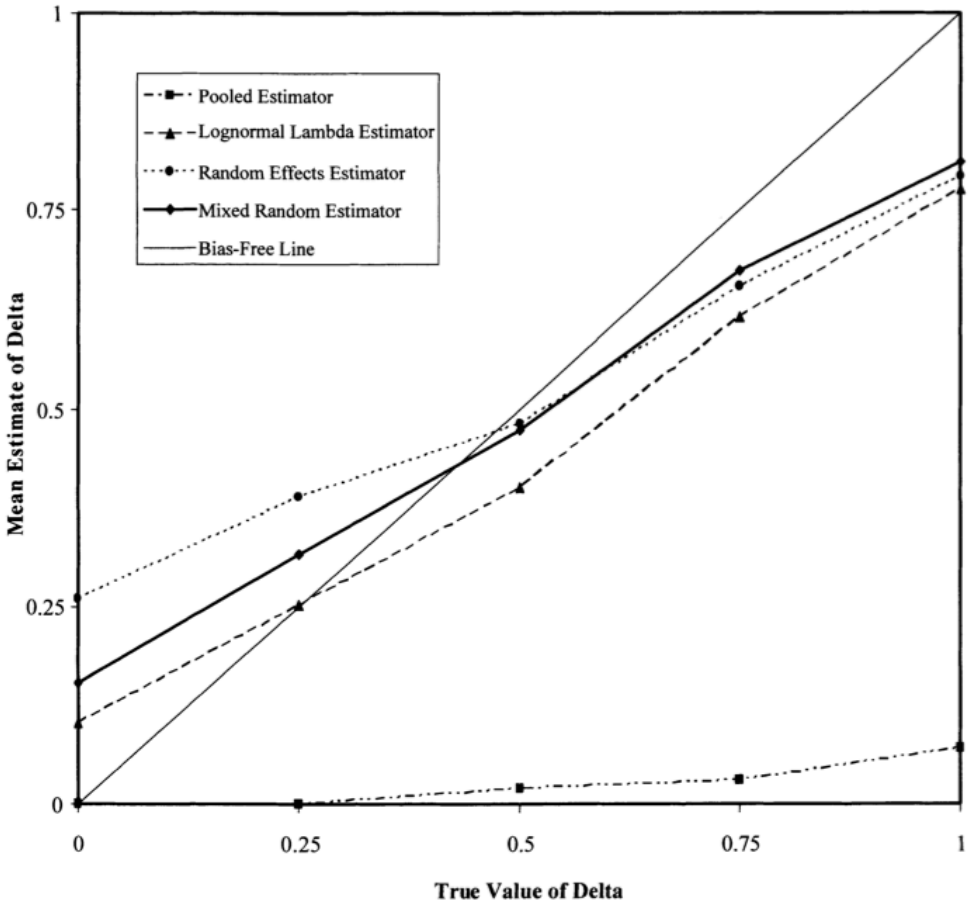

Figure 2: Individual heterogeneity in learning would lead to severely biased estimations of delta in pooled estimation. The graph shows the estimated value of delta with a number of different estimation schemes based on simulated learning behavior with high individual heterogeneity. As shown in the graph, when true delta is 0.5, the mean of estimated delta based on pooled estimation is around 0 indicating the pool of the subjects are reinforcement learners whereas in fact they are made of equal share of reinforcement and belief based learners. Adapted from Wilcox (2006).

Some economists (Gul and Pesendorfer 2008; Gul and Pesendorfer 2009) may argue that if two economic models both fit behavioral data well enough, there is no need to distinguish between them, as any economic model is after all an approximation to the true underlying model. Such a claim is certainly valid from neoclassic economics point of view. However, if we care for out of sample prediction, such as to predict how agents

will behave in an unfamiliar game setting, how policy changes will affect behavior, or how another population will respond to the same external settings, we do need to have a better understanding about the underlying learning process involved under different strategic settings. This idea is summarized succinctly in Salmon (2001).

> *"If the goal of the research is to find a model that has a high in-sample prediction level, then the problem discussed here is not an issue. If the point, however, is to further our understanding of strategic choice, then this is a more serious problem. This is why there will be so much emphasis placed on accuracy of the models in identifying the true data generating process instead of simply providing statistically acceptable fits of the data."*

Here we provide an alternative approach for identifying the underlying strategic learning process, a neuroeconomic approach. Our goal is to relate the game theoretical models to neural measures, in order to test competing hypotheses that constrain competing learning algorithms, and to further our understanding in strategic choices.

1.2. A neuroeconomic approach to learning

Evidences for learning in neuroscience

The intimate relationship between brain and reward has been appreciated by neuroscientists since the accidental discovery of the mid-brain dopamine neurons in the 1950s. In the seminal work by Olds and Milner (1954), they found that rats would repeatedly visit the locations where they received direct electrical stimulation of mid-brain regions dense in dopamine neurons. It is also found by later studies that when rats

were allowed to press a foot lever to stimulate this particular brain region, they would do so in preference to food, drink, or mate. Conversely, blocking dopamine has the effect of removing the reward contingent on an animal's choices. For example, Berridge (1998) demonstrated the extinction of responding to food reward as if the reward is absent, when animal is deprived of dopamine.

Over the past decades, researchers have further elucidated the role of this brain region in decision-making. A series of studies by Schultz and colleagues (1997) were particularly influential in moving beyond the relatively naïve hypothesis of reward encoding in mid-brain dopamine neurons. These single neuron recordings in monkeys used a classical conditioning task where monkeys learned to associate a stimulus (e.g., light) with a reward (e.g., juice). Contrary to the simple reward encoding hypothesis, the dopaminergic neurons fired in response to reward delivery only in early trials. Once the monkey learned the association between the light and juice, activity in the neurons shifted to the onset of the light. Moreover, the firing rate of the neurons fell below baseline if the light was followed by the omission of the juice reward (Figure 3).

The TD reinforcement model provides a flexible framework for studying the brain mechanisms underlying a variety of conditions. Over the past decade, it has proved remarkable success in its application to the study of neural systems underlying decision-making and learning. In particular, it forms one of the foundational results in the rapid growing field of neuroeconomics, and has been found to be robust across a range of species including humans and nonhuman primates as well as experimental conditions (Fiorillo, Tobler et al. 2003; Padoa-Schioppa and Assad 2006; Rangel, Camerer et al. 2008). For example, Figure 4 shows the result of an early study of brain regions that

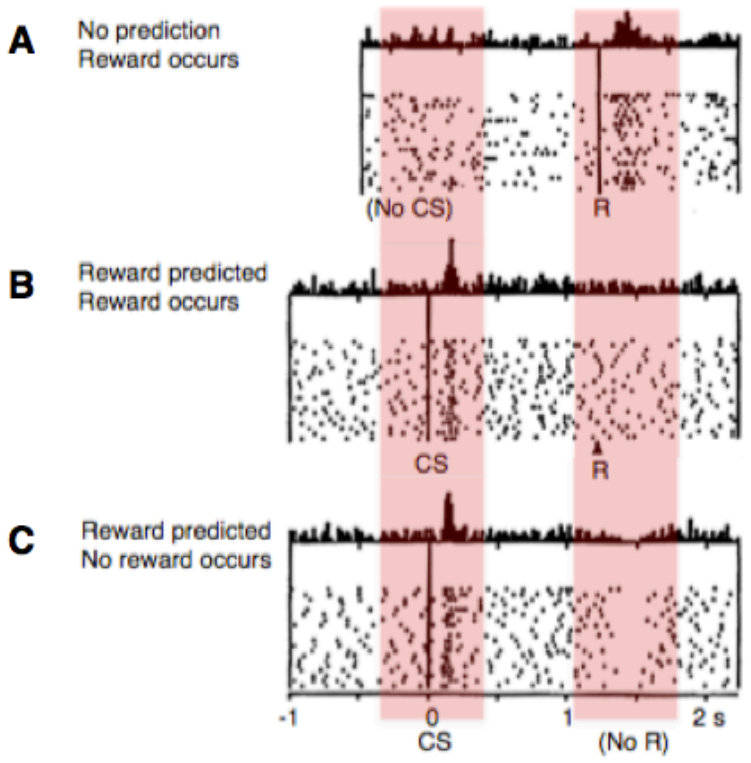

Figure 3: Changes in dopaminergic neurons resembles reward prediction errors when monkeys performing a conditioning task to associate a stimulus with a reward. Raster plots demonstrate dopamine neuron's activity with each row representing a trial, aligned to the time of conditioned stimulus (CS) or reward (R/ No R). Histograms show the summed activities over each trial. (A) Dopaminergic neurons respond to the unexpected delivery of juice with a phasic spike of firing. (B) After conditioning the delivery of juice with the predictive light stimulus, reward is delivered as predicted, no reward prediction error occurs. The dopamine neuron is activated only by the reward-predicting stimulus but not by the reward delivery. (C) When reward is unexpected omitted, prediction error occurs. Dopaminergic neurons showed a pause in firing, below their standard background firing rate. Adapted from Schultz, Dayan et al. (1997)

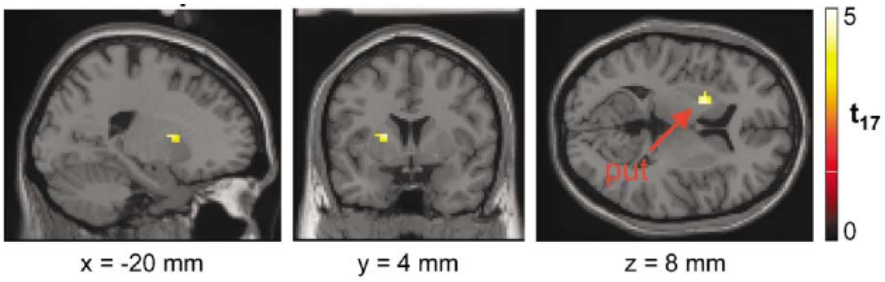

Figure 4: Human ventral striatum (in particular, left putamen) responds to positive TD prediction errors caused by unpredictable juice delivery in a passive learning task. Adapted from McClure, Berns et al. (2003)

responds to the TD prediction errors in a passive learning task. In this study, McClure, et al. (2003) scanned human subjects while they underwent a classical conditioning paradigm in which associations were learned between visual stimuli (a flash of light) and the time of reward delivery (juice). Subjects were then exposed to low frequency of "error" trials, in which the delivery of reward is delayed beyond the time expected from the previous training session. They found that unpredicted delivery was associated with greater activities in some brain region (ventral striatum) than predicted reward delivery, which suggests that computations described by TD learning are expressed in the human brain during this learning task.

Neuroeconomics at a glance

Neuroeconomics is an interdisciplinary field of economics, neuroscience and psychology, which combines neural measurement technology, economic theories, experimental paradigms, as well as mathematical modeling. The field concerns with grounding microeconomic models and theories with neural measures, in order to build hypotheses that constrain alternative economic theories, and thus provide at a deeper level the empirical evaluations and distinctions among competing behavioral models as

well as empirical inspiration for economic theories to incorporate the more realistic psychological and neurological underpinnings.

The underlying assumption in neuroeconomics is that behavioral choices are the results of certain information processing which involves brain activities. Hence neural signals can be used to test economic theories on what information is processed in the brain, how such processing takes place and how choices are decided (Sanfey, Loewenstein et al. 2006; Fehr and Camerer 2007; Camerer 2008; Harrison 2008; Krueger, Grafman et al. 2008; Bernheim 2009; Sobel 2009).

To illustrate, consider a decision problem in a repeated matching penny game. Traditional economic models seek to identify the casual relation $f(y \mid x,u)$ of individual choices y (e.g. selecting head or tail) conditional on a vector of observed environmental variables x (e.g. payoff structure, observed history of previous choices) as well as latent variable u (e.g. belief about what the other player will do) under some axioms.

To illustrate, consider an individual decision-making problems. Traditional economic models seek to identify the causal relation $f(y \mid x)$ of individual choices y (e.g. purchasing behavior) conditional on a vector of observed environmental variables x (e.g. prices, incomes etc.) under some axioms. Whereas unobserved variables such as moods, traits, memories, are treated as idiosyncratic noises orthogonal to the observed explanatory variable x. In contrast, neuroeconomics seeks to open the black box of the decision making process $f(y \mid x)$ (W. Glimcher, Camerer et al. 2008), by assuming there exists a computational decision algorithm at the brain level that takes external stimuli x as inputs and generates behavioral choices y as outputs, while at the same time producing neural signals z that is observable under neuroscientific methods. i.e. $g(z,z' \mid x)$ and

$g'(c|z,z')$, where z and z' are observed and unobserved neural signals respectively. This approach allows economists to derive testable hypotheses based on both neural and behavioral data, and conduct more stringent test on the underlying economic theories (Rustichini 2009). which has been proven useful in clarifying some long standing behavioral paradoxes (Hsu, Bhatt et al. 2005), or adjudicating among different economic models which can explain the same pattern of choices (Knoch, Pascual-Leone et al. 2006).

1.3. Model-Based Integration of Behavioral, Neural, and Lesion Data

My thesis aims to take an initial step toward providing such a framework by bringing together a group of experts in (1) the neuroeconomics of social and non-social decision-making (Hsu et al. 2005; Hsu et al. 2008), (2) neural modeling (Duque et al. 2010), and (3) lesion patient studies of cognition and executive function (Barceló and Knight 2002; Beer et al. 2006). In particular, we will extend the RL framework and provide a model-based account of two fundamental components of goal-directed behavior in the social domain: (1) learning about the rewards and punishments available in the environment (the reward learning problem), and (2) forming mental models of other individuals in the decision context (the theory of mind problem). In this way, we seek to provide a unifying account of goal-directed behavior in both social and non-social settings.

Integration of Modeling and Experiments

The overarching innovation of this study is the tight integration of computational modeling encompassing goal-directed behavior in both social and non-social settings, and testing the predictions generated using complementary experimental techniques. Experimentally, we will use fMRI of healthy adults and behavioral studies of focal brain lesion patients to triangulate the neural computations and the causal mechanisms. Bridging the different levels of explanation will be the computational model that captures the internal representations of the quantities, or "latent values", that drive behavior, and will be used in a number of different ways. First, we will use the model, calibrated on observed behavior, to derive trial-by-trial regressors for use in our fMRI experiments. Second, the estimated parameters of the model themselves can be used to compare across health and diseased groups, or find subtypes of the diseased groups. Finally, the neural correlates and the behavioral estimates can be combined in order to find novel brain-behavior markers of diseases.

This approach has become widespread in certain areas, including much of reward learning (Daw et al. 2006; O'Doherty et al. 2007) and parts of impulsivity (Kable and Glimcher 2007; Balsam and Gallistel 2009) and risky decision-making (Tobler et al. 2007; Hsu et al. 2009). It is comparatively rare in the social domain. The twin challenges here are generating models of sufficient predictive power and tractability to describe fundamental components of social behavior, and to have sufficient neurobiological plausibility that the putative internal quantities reflect biological reality, which is particularly challenging for fMRI studies (Behrens, Hunt et al. 2008; Behrens, Hunt et al. 2009); Hsu et al. 2008).

Connecting RL and Game Theory

The two core innovations that make our approach possible are: (1) connecting existing theoretical frameworks of reinforcement learning and behavioral game theory, and (2) development of experimental paradigms that taps into goal-directed behavior in the social domain and, at the same time, can be used across different experimental methodologies. The first feature allows us to leverage the behavioral realism of the behavioral game theory framework by building on top of the neurobiological realism of the RL framework. Specifically, we adapted behavioral models of social learning into a temporal-difference framework proposed for reward learning and decision-making (Rangel, Camerer et al. 2008). This is the critical step in generating trial-by-trial regressors necessary for model-based fMRI studies. We further verified the validity of the model with model-free measures of behavior (e.g., switch/stay). This gives a qualitative look at the data that serves as a "sanity check" of the model predictions.

Novel Behavioral Paradigm

We developed the second component by combining separate paradigms used in reward learning and social neuroscience, respectively. Subjects will be presented with a choice paradigm—a game—where outcomes are determined by both the choices of the subject and other(s) in the game. This is done using a novel multi-strategy competitive game that is amenable for computational modeling and optimized for statistical recovery of key parameters of the model. Furthermore, we adapted the paradigm to work equally well in a neuroimaging and lesion experimental setting by minimizing the cognitive

complexity of the interface that often accompany game theory paradigms. Specifically, we replaced the large matrix of numbers used in typical game theory studies, which can be unintuitive to even highly educated healthy subjects, with an interface that directly reflects the logic of the game. To manipulate the social context of the subjects' behavior, we will borrow from the social neuroscience and neuroeconomics literature and vary the other players in the game between human and computer opponents (McCabe et al. 2001; Gallagher and Frith 2003). Furthermore, this paradigm may be further extended by incorporating features that have been profitably studied in the existing literature. This includes manipulating the moral status of the opponent (Delgado, Schotter et al. 2008), matching subjects against a single opponent to allow reputation building (King-Casas et al. 2005; Kishida et al. 2010), or developing computer algorithms that mimic human behavior.

Bridging Existing Approaches to Goal-Directed Behavior

Because we start with simple forms of social behavior that are tractable for both modeling and experimentation, our bottom-up approach can serve as a potential bridge between the large and comparatively well-developed literatures of reward learning on the one hand, and social exchange games on the other. The latter includes the highly popular games such as Trust (Delgado et al. 2005; King-Casas et al. 2005) and Ultimatum Games (Sanfey, Rilling et al. 2003). Indeed it was work in this tradition that first provided empirical support for the idea that social interaction can operate through a RL mechanisms as more basic types of rewards (King-Casas et al. 2005). However, because of the complexity of behavior arising from these games, and the substantial individual

differences, it is extremely challenging to develop models that offer a mechanistic understanding of the brain-behavior link. In contrast, drawing from both theoretical biology and game theory (Fudenberg and Levine 1998; Hofbauer and Sigmund 1998), we will begin with perhaps the best-studied and most well-understood theoretical scenario for social interactions containing simple forms of mentalization. We envision that by alternatively stripping away and "scaling up" the social complexity of our experimental paradigm, we will be able to connect and interface directly the theoretical and experimental frameworks of both literatures.

Finally, we should note that we by no means claim that this framework encompasses all, or even large proportion of social behavior and functioning. Rather, we argue that such a framework has to potential to do for goal- directed behavior in the social domain what it has demonstrated in the non-social counterpart. Given the limited state of our knowledge, even the small step we propose here constitutes a productive and substantial advance.

CHAPTER 2

NEURAL CORRELATES UNDERLYING STRATEGIC LEARNING

2.1 Introduction

Notions of equilibrium are central to theories of strategic behavior in the biological and social sciences where individuals must take into account not only available rewards and punishments but also behavior of other intelligent agents. It has long been recognized, however, that equilibria do not emerge spontaneously, but rather through some adaptive process whereby organisms evolve or learn over time (Fudenberg and K. Levine 1998; Hofbauer and Sigmund 1998). In the social domain, specifically behavioral game theory, this is referred to as strategic learning, and has been the subject of intense study both theoretically and empirically. Only recently however are we beginning to understand the neural mechanisms and computations involved (Barraclough, Conroy et al. 2004; Hampton, Bossaerts et al. 2008; Lee 2008). Here we study the neural systems underlying strategic learning in a stylized but well-characterized setting of a population with many anonymously interacting agents, which provides a natural model for settings such as commuters in traffic or bargaining in bazaars (Fudenberg and K. Levine 1998). More importantly, it minimizes the role of reputation and higher-order belief considerations, and has served as a basic building block for a number of models in evolutionary biology and game theory (Fudenberg and K. Levine 1998; Hofbauer and Sigmund 1998).

We considered three classes of models commonly invoked to explain behavior in this population setting—reinforcement learning (RL), belief-based (BB) learning, and their hybrid, experience-weighted attraction (EWA) (S. Sutton and G. Barto 1998; Camerer 2010). RL models, which seek to understand how choices are shaped by rewards and punishments, were among the first to be applied to strategic learning, and explain a number of important features of the data at the behavioral (Mookherjee and Sopher 1994; Roth and Erev 1995; Mookherjee and Sopher 1997) and neural levels (Barraclough, Conroy et al. 2004; Dorris and Glimcher 2004). This approach is particularly appealing as there is much evidence of a temporal difference (TD) form of RL at the neural level, whereby learning is driven by a reward prediction error defined as the difference between expected and received reward (Schultz, Dayan et al. 1997; Rangel, Camerer et al. 2008). Despite these successes, standard RL models provide an incomplete account of strategic learning even in the simple population setting, as they have no notion of the actions and beliefs of other agents and can be consequently exploited in a variety of competitive settings (Camerer 2010). In contrast, BB models require players to form and update first-order beliefs regarding the likelihood of future actions of opponents through experience. Specifically, these models posit that players select their actions strategically by best responding to their beliefs about opponents' future strategies. Although belief formation can take any number of forms, such models typically assume that players use some weighted history of opponents' choices as basis for their beliefs (Camerer and Ho 1999).

In principle, reinforcement and belief learning are not mutually exclusive. Hybrid models have been proposed to combine RL and BB models. Here we adopt EWA, a nonlinear hybrid of RL and BB models that nests both (Camerer and Ho 1999). At the

behavioral level, the hybrid model has been found to outperform both standard RL and BB across a wide range of games (Camerer and Ho 1999; Ho, Camerer et al. 2007). However, it is unclear whether the behavioral superiority is the result of a single learning signal at the neural level incorporating both reinforcement and belief inputs, or an interaction of dissociable signals. Recent work in related fields—primarily reward learning and social cognitive neuroscience—provide converging evidence for distinct systems that underlie reinforcement processing, compared to those that represent mental models of the environment or other intelligent agents (Amodio, 2006; Behrens, Hunt et al. 2009). Based on these findings, we hypothesize that there exists distinct neuronal populations encoding RL and BB prediction errors, and that both can be represented under a TD framework.

The rest of the paper is structured as follows. Section 2.2 gives a brief background introduction including the quantitative learning models, and the experimental paradigm we borrowed from game theory. Then in Section 2.3 we introduce experimental and computational methods employed in the study. Section 2.4 presents the major behavioral and section 2.5 presents the neuroimaging results. We conclude with a discussion of our results and their relevance to the various disciplines in Section 2.6.

2.2 Computational Modeling of Strategic Learning

I consider here three classes computational models of strategic learning: (1) reinforcement learning, (2) belief-based learning and (3) EWA learning. All three models assume that a decision-maker chooses an action from a pool of actions based on some subjective value assigned, and that these values are updated on each trial according to

some prediction errors. They differ, however, in the hypotheses of how prediction errors are computed over the course of learning.

Reinforcement learning hypothesizes that prediction errors are computed based on agents' received payoffs (McClure, Berns et al. 2003; O'Doherty, Critchley et al. 2003; Daw, O'Doherty et al. 2006). More specifically, it assumes that actions that yielded the most reward in the recent past will receive higher subjective valuation and hence be selected with higher chance in the future. In contrast to reinforcement learning, belief-based learning posits that individuals form beliefs about the behavior of other players, and then maximize the subjective value by best responding to these beliefs. Because beliefs are not directly observable, these models typically make assumptions about the mapping from history of opponent play to beliefs (e.g., Fudenberg and K. Levine 1998). For example, the well-known fictitious play model posits that players form beliefs about the opponent based on the average of the opponent's history of play (Fudenberg and Kreps 1993).

The third model we consider is EWA (Camerer and Ho 1999; Ho, Camerer et al. 2007). A key insight of EWA is the recasting of belief-based learning into reinforcement learning using foregone payoffs. This is not only theoretically appealing but also biologically plausible, as simple forms of fictive errors have been found in both human and non-human primate studies of reward learning (Lohrenz, McCabe et al. 2007; Hayden, Pearson et al. 2009; Thevarajah, Mikulic et al. 2009). EWA has the advantages of (1) providing a general framework that contains both reinforcement and belief-based learning as special cases, and (2) allowing for a more sophisticated way to account for the

depreciation of the past by separating the discount of the past experience and the discount of the past subjective values.

For example, in EWA, we denote the values as the "attraction" for player i to play strategy k, after observing the previous t rounds of play, i.e., $A_i^k(t)$. The evolution of the attractions is governed by three parameters. Two of the parameters, ϕ and ρ, control the depreciation of $A_i^k(t)$ and the strength of past experiences, captured by $N(t)$. For example, if the player believes his opponent is a fast adaptor, he will adapt fast too. That is, he will have a small ϕ that depreciates past attractions faster. In contrast, ρ is the discount rate for the strength of past experience $N(t)$, and hence controls the influence of the initial attractions. If ρ is large, the initial attractions will wear off quickly. Finally, parameter δ is the weight between foregone payoffs and actual payoffs when updating attractions. This parameter can be interpreted as a psychological inclination toward foregone payoff learning, which is equivalent to belief-based learning (Camerer and Ho 1999).

Begin with initial attractions over strategies $A_i^k(0)$ and initial experience $N(0)$, which can either be estimated as free parameters or calculated from first period data (Roth and Erev 1995; Saxe and Haushofer 2008), $A_i^k(t)$ and $N(t)$ evolve according to the following equations:

$$A_i^k(t) = \begin{cases} \dfrac{\phi \cdot N(t-1) \cdot A_i^k(t-1) + \pi_i\left(s_i^k, s_{-i}(t)\right)}{N(t)}, & \text{if } s_i^k = s_i(t) \\ \dfrac{\phi \cdot N(t-1) \cdot A_i^k(t-1) + \delta \cdot \pi_i\left(s_i^k, s_{-i}(t)\right)}{N(t)}, & \text{if } s_i^k \neq s_i(t) \end{cases}$$

$$N(t) = \rho \cdot N(t-1) + 1, \quad t \geq 1 \ .$$

Attractions can be further rewritten a form conducive for computational estimation by rewriting in the form of:

$$A_i^k(t) = \frac{\phi \cdot N(t-1) \cdot A_i^k(t-1) + \left[\delta + (1-\delta) \cdot I(s_i^k, s_i(t))\right] \cdot \pi_i(s_i^k, s_{-i}(t))}{N(t)}$$

where $I(s_i^k, s_i(t))$ is equal to 1 if and only if player i played strategy k in the previous round t. Otherwise, $I(s_i^k, s_i(t))$ is equal to zero. When $I(s_i^k, s_i(t))$ takes the value 1 the actual payoff is updated, whereas when it's equal to 0 the fictive payoff is updated.

I also estimated reinforcement and belief-based learning models as two restricted versions of the above EWA model. When $\delta = 0$ and $N(0) = 1$, EWA reduces to the reinforcement model, similarly when $\delta = 1$ and $\rho = \phi$, it reduces to the belief-based learning model.

To convert attractions into choices, we use a logit or softmax function to calculate the probability of player i playing strategy k in the next rounds, i.e.

$$p_i^k(t+1) = \frac{e^{\lambda A_i^k(t)}}{\sum_{l=1}^{L} e^{\lambda A_i^l(t)}},$$

where λ is a measurement of subjects' sensitivity to attractions.

Statistical Properties of Experimental Paradigm

Statistically, strategic learning models, particularly EWA, place high demands on the quality and quantity of data. Previous econometric studies have found that games with many choices have superior statistical properties than smaller games (Salmon 2001), which appear borne out in experimental studies (Erev and Rapoport 1998; Feltovich 2000). This is notable in that the few previous studies of neural basis of strategic learning, in both humans and non-human animals, have used 2x2 games, the smallest

non-trivial games (Dorris and Glimcher 2004; Hampton, Bossaerts et al. 2008; Lee 2008). Besides the consideration from statistical estimation point of view, the foregone payoff and the received payoff are highly negatively correlated in such small games with only two possible outcomes given what opponent has selected. This makes it impossible to separate the signals associated with foregone payoffs and realized payoffs.

Games with many strategies and possible outcomes, however, can take a long time to play and to process cognitively. In the experimental economics literature on learning, sessions exceeding an hour are typical, and up to two hours are not unusual (Rapoport and Amaldoss 2000; Camerer, Ho et al. 2003). Such durations are clearly unfeasible in the context of neuroimaging studies. The solutions are: (1) to acquire data for a fewer number of rounds, or (2) complete part of the behavioral experiment outside the scanner. Neither is ideal. Another serious concern is the extended duration of the decisions themselves, often-exceeding 10 sec, in some cases minutes. This makes it difficult to implement event-related analyses of the functional data. We believe that these

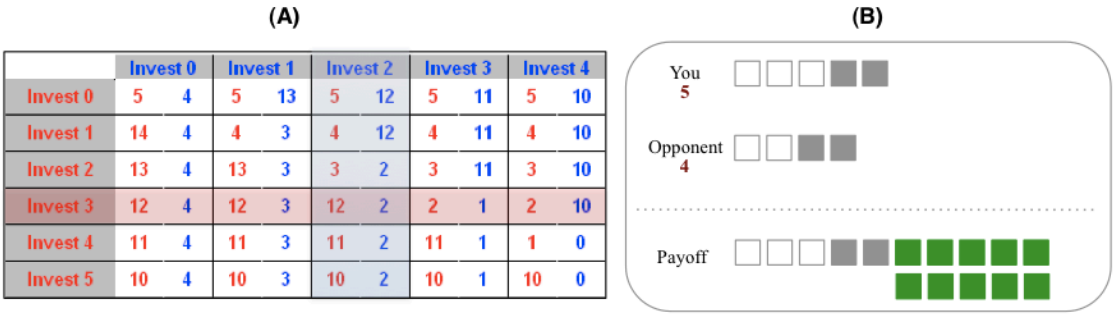

Figure 5: The patent race game. Players are given an endowment to invest (gray squares). Player who invests more receives the potential prize (green squares) plus portion of her endowment not invested. Player who invests less receives only the portion of her endowment not invested. In case of tie neither player receives the prize. Compare (A) the standard representation of the patent race game with (B) our alternative.

factors signals in previous studies (Lee, McGreevy et al. 2005; Hampton, Bossaerts et al. 2008). In contrast, we show in our preliminary data strong evidence for the existence of such regions.

To overcome these challenges, we used the "patent race game", first studied experimentally by Rapoport and Amaldoss (2000). This game is simple in motivation but rich in the strategic nuances and the patterns of behavior that it can generate. In the game, two players are randomly matched from a large pool of players at the beginning of each round and compete for a prize by choosing an investment (in integer amounts) from their respective endowments. The player who invests more wins the prize and the other loses. In the event of a tie, both lose the prize. Players keep the part of their endowment that is not invested. In the particular payoff structure that we use, the prize size is 10, and players are of two types: *Strong* and *Weak*. The Strong player has 5 units to invest, resulting in the strategy set $\{0,1,2,3,4,5\}$, whereas the Weak player has 4 units to invest, with a resulting strategy set of $\{0,1,2,3,4\}$ (Figure 5). The large strategy space of this 6x5 game improves the recovery of key parameters in learning models (Salmon 2001). Theoretical game theory predicts that the game has a unique mixed strategy Nash equilibrium in which strong players invest five 60% of the time, one and three 20% of the time respectively, and weak players invest zero 60% of the time, two or four 20% of the time.

2.3 Methods

Experimental Design

We redesign the interface of the patent race game by taking the standard matrix form representation that contains 60 elements and replacing it with a display that directly reflects the logic of the game. This speeded up game play considerably. For example, subjects took over 2 hours to run through 160 rounds of this particular game (Rapoport and Amaldoss 2000). In contrast, our preliminary data show that subjects were able to complete 160 rounds in around 30 minutes. Including typical inter-stimulus intervals (ISI) values of fMRI experiments, almost all subjects were able to complete the scanning session in 40 minutes.

The experiment contains two sessions, a behavioral session and a neuroimaging session. 16 subjects participated the behavioral session. In the beginning of each round subjects are randomly and anonymously matched. They played 2 sessions of 80 rounds each to counterbalance the ordering effect. In the neuroimaging session, however, the opponent choices are drawn from a pool of the original 16 subjects who participated in a behavioral experiment (we will refer to these as pool players). These neuroimaging subjects played in the same sequence as pool players. That is, if the scanner subject is playing in round 60, the opponent's choice will be drawn randomly from round 60 of one of the pool player. Similar to the behavioral session, neuroimaging subjects also played 2 sessions, each of 80 rounds.

This is done for two reasons. First, the random matching protocol requires a large number of subjects to ensure that the probability of repeated interaction is small. Otherwise, subjects may be able to develop higher order strategies that depend on

reputation, which complicates the analysis considerably. Second, random matching makes running real-time experiments with opponents impractical. For example, over 200 subjects will be used to acquire 20 neuroimaging sessions with 10 players, a relatively small group. Our current method preserves the dynamics of the evolution of play in the experiment, as well as controls for the inter-group variation that would arise if we used more than one group of pool subjects.

The timeline of experiment is as follows. After a fixation screen of a random duration, distributed uniformly random between 4-8s, subjects are presented with the patent race game for between 1-2s (Figure 6A). Subjects are allowed to make a decision once the dashed line turns from red to gray (Figure 6B). Subjects input the decision by pressing a button mapped to the desired investment amount. The possible investments are {0,1,2,3,4,5} for the strong player and {0,1,2,3,4} for weak. All decisions are self-paced. After 2-6s, if the investment amount is strictly greater than that of the opponent, the subject wins the prize (Figure 6C). Otherwise the subject loses the prize (Figure 6D). In either case the subject keeps the portion of the endowment that was not invested.

Subjects

A total of 35 healthy volunteers (19 female) were recruited from Neuroeconomics Lab subject pool at the University of Illinois at Urbana-Champaign (UIUC). Informed consent was obtained as approved by the Internal Review Board at UIUC. Subjects had a mean age: 23.3 +/- 4.6 years, ranging 19-47. 5 subjects were excluded from the study due to excessive motion, and 3 due to repeating the same strategy for more than 95% of the trials during the experiment.

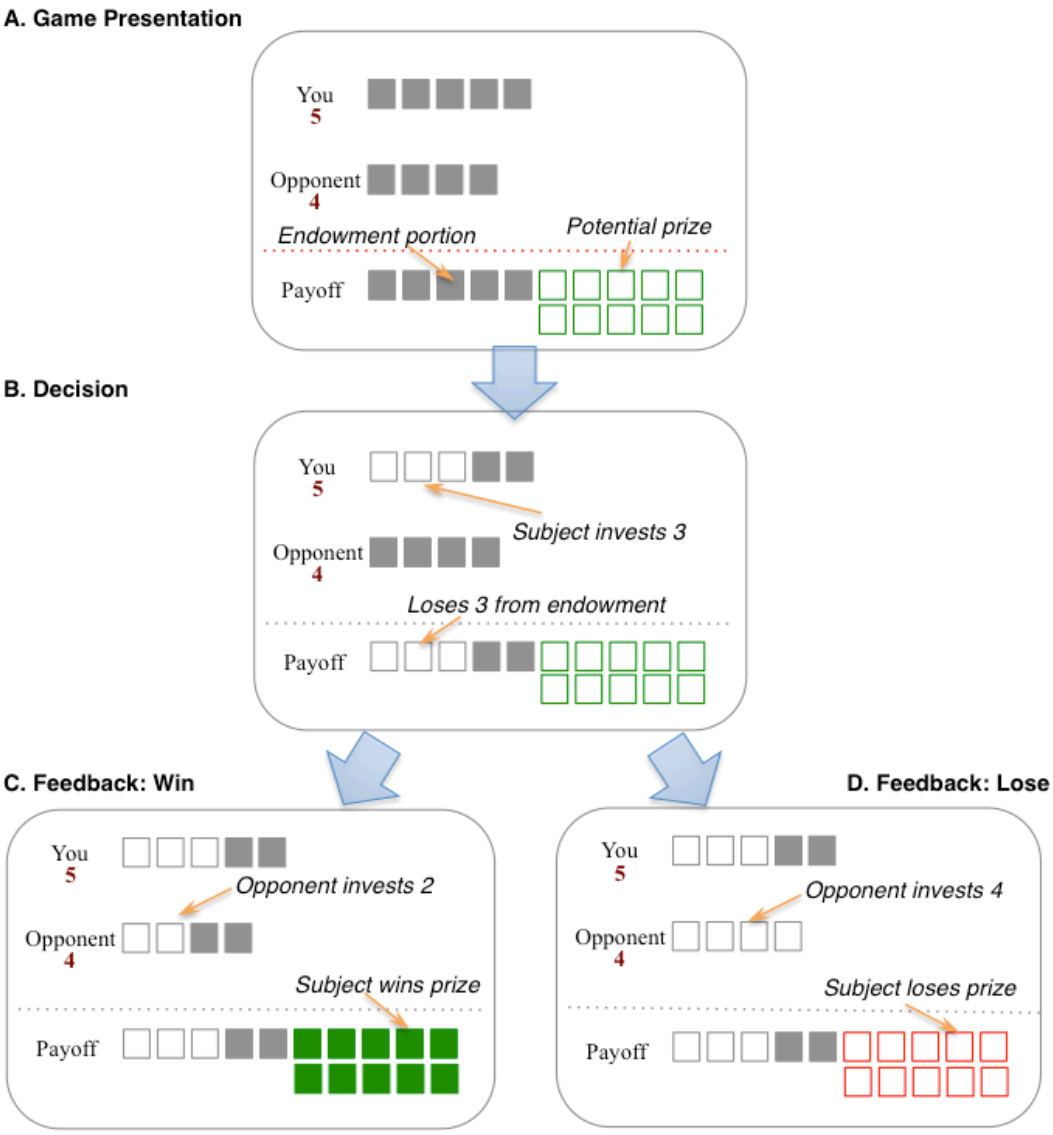

Figure 6: Timeline of experiment. After a fixation screen of between 4-8 sec, (A) the subject is presented with the patent race game for between 1-2sec, (B) the subject is able to make a self-paced decision once the dashed line turns from red to gray by pressing buttons mapped to the investment amounts. After 2-6sec, the opponent's choice is revealed. (C) If the investment amount is strictly greater than that of the opponent, the subject wins the prize. (D) Otherwise the subject loses the prize. In either case the subject keeps the portion of the endowment that was not invested.

Before entering the scanner, subjects were given instructions and completed a quiz to ensure comprehension of the game. They were informed that they would be paid the average payoff from a randomly chosen 40 rounds from each session, plus a $10 show up fee.

Estimation Method

To evaluate how closely the three learning models capture subjects' behavior, we fit each model to behavior separately. In particular, we perform both pooled and individual-level estimation. For pooled estimation, we aggregate observations from all subjects conditional on their roles and then fit the choice data by maximizing the log likelihood of the observed choices over rounds for subject i. That is, $\sum_i \sum_t \log\left(p_i^{s_i(t)}(t)\right)$. To estimate the model, we conduct a grid search over a large range of values for all free parameters, since the likelihood function is not globally concave.

The pooled estimation essentially assumes that a single, shared set of parameters can explain the behavior from all subjects' with the same endowment. Although pooled estimation is more robust in general, it removes the individual differences in learning that is important for between subject analysis in neuroimaging and may bias behavioral estimates due to heterogeneity (Wilcox 2006). Therefore, we investigate fully individualized fits with separate parameters for each subject. The primary challenge of individual-level estimation is the relatively small sample size compared to the number of free parameters. We approach this problem with two methods. First we apply the self-tuning estimation as introduced in Ho et al. (2007). The psychologically less interesting parameters such as $N(0)$, ρ and initial attractions are set to be fixed values. ϕ is

replaced by a function of past experience as in Ho et al. (2007). Different from self-tuning EWA, however, δ is kept as a free parameter in our estimation, because it captures the individual differences in weighting the respective learning processes critical to the neuroimaging analysis. Second, we also conduct partially joint estimates across different roles for each subject. These methods appeared quite robust and yielded similar results.

Conversion to Temporal-Difference Model

Having characterized subjects' behavior computationally, we use the best-fitting estimates of each model to generate regressors containing prediction errors for each subject on each trial. This requires us to convert the standard quantitative learning models into the TD forms, where reward predictions and prediction errors are separated into different components. TD models have been highly popular in computational neuroscience, and forms the bedrock of much of our current understanding of how the brain learns. EWA can be transformed into the following forms for updating attractions under the restriction of $\rho = \phi$. This is a mild restriction, and it has been shown to have little effect on empirical fit across a number of datasets (Ho, Camerer and Chong 2007).

$$A_i^k(t) = \underbrace{A_i^k(t-1)}_{\text{Reward Prediction}} + \underbrace{\begin{cases} \frac{1}{N(t)}\left\{\pi\left(s_i^k(t), s_{-i}(t)\right) - A^k(t-1)\right\} & \text{if } s_i^k = s_i(t) \\ \frac{1}{N(t)}\left\{\delta \cdot \pi\left(s_i^k(t), s_{-i}(t)\right) - A^k(t-1)\right\} & \text{if } s_i^k \neq s_i(t) \end{cases}}_{\text{Prediction Error}}$$

In words, if player i played strategy k in the previous round, the new EWA attraction $A_i^k(t)$ is the previous attraction plus the discounted prediction error. If strategy k was *not* played in the previous round, the new EWA attraction $A_i^k(t)$ is the previous

attraction plus the discounted prediction error multiplied by the foregone payoff parameter. In contrast, in the standard TD model the valuation of unchosen strategy k is assumed to remain the same or decay at some depreciation rate. This is clear in our reinforcement learning TD model below. Finally, we include the belief-based model equation. The intuition here is similar to the EWA model, except all prediction errors are weighted equally.

$$\text{RL:} \qquad A_i^k(t) = A_i^k(t-1) + (1-\phi)\left\{\frac{1}{1-\phi}\pi\left(s_i^k(t), s_{-i}(t)\right) - A^k(t-1)\right\}$$

$$\text{BB:} \qquad A_i^k(t) = A_i^k(t-1) + \frac{1}{N(t)}\left\{\pi\left(s_i^k(t), s_{-i}(t)\right) - A^k(t-1)\right\}.$$

fMRI Data Acquisition

Functional MR images were obtained for each subject using a 3T Siemens Allegra scanner located at the research-dedicated Beckman Imaging Center (BIC) at the University of Illinois at Urbana-Champaign. Images were acquired using echo-planar T2* images with BOLD (blood oxygenation-level-dependent) contrast, and angled 30 degrees with respect to the AC-PC line to minimize susceptibility artifacts in the OFC. MR imaging settings are as follows: repetition time (TR) = 2000ms; echo time (TE) = 40ms; slice thickness = 3mm yielding in a 64x64x32 matrix (3mm x 3mm x 3mm); flip angle = 90 degs; FOV read = 220mm; FOV phase = 100mm, series order: interleaved. High-resolution structural T1-weighted scans (1mm x 1mm x 1mm) were acquired using a MPRage sequence. Visual stimuli were viewable by means of a mirror mounted on the MRI head coil and responses will be acquired via an MRI-safe button response pad (Neuroscan, Inc.).

fMRI Data Analysis

Imaging data was preprocessed using *SPM2*, and included, in order, slice time correction, motion correction, coregistration, normalization to the MNI template and smoothing of the functional data with an 8mm kernel. Random effects analyses were done in *SPM2* by specifying a separate general linear model for each subject and pooled at the second level. First all images were high-pass filtered in the temporal domain (filter width 128s) and autocorrelation of the hemodynamic responses was modeled as an $AR(1)$ process.

Functional imaging data was analyzed using standard GLM procedures. In our GLM model, each trial is modeled as two events: the decision event and the feedback event. Regressors are constructed by using the trial-by-trial output from the TD form of the best-fit behavioral models. The decision event is associated with choice probabilities, which can be regarded as relative reward predictions controlled for time influence. The feedback event is associated with prediction errors for chosen actions. Regressors were then convolved with the canonical hemodynamic response function and entered into a regression analysis against each subject's fMRI data using SPM 2. The regression fits of each computational signal from each individual subject were taken into random-effects group analysis.

More specifically, for the TD transformation, prediction error for reinforcement learning model is $(\frac{1}{1-\phi}\pi_{it} - v_{i(t-1)})$, whereas for the EWA model it is $(\pi_{it} - v_{i(t-1)})$. For belief-based learning model, however, prediction error is defined slightly differently. Instead of using the absolute prediction errors $(\pi_{it} - v_{i(t-1)})$, we find it is crucial to use the prediction error for the chosen action relative to the fictive errors in the same trial. The

reason is the following: the conventional TD model is based on the action-outcome evaluation system, which focuses on reinforcing the associated outcome for each action. The core idea for belief-based learning model, however, is that an action should not be only evaluated by its directly associated outcome but also by the fictive outcomes conditional on opponent's choices. Hence absolute prediction error fails to capture the essence of belief-based learning model. For example, in our game the strong player is guaranteed with 10 units of payoff by investing 5 regardless of his opponent's investments. As a result his absolute prediction error for investing 5 will always be 0. However, he might regret for investing 5 when observing his opponent invested 0, and be relieved if the opponent invested 4. In other words, the relative prediction error for investing 5 when the opponent invests 0 is lower, which will in turn reduce the likelihood of him investing 5 in the future.

A few forms of relative prediction errors are considered in this study, including orthogonalized prediction errors for chosen actions against the maximum fictive error in the same trial, orthogonalized prediction errors for chosen action against mean prediction errors for all actions in the same trial, the exponential ratio between prediction error for the chosen action and the mean prediction errors for all actions $\frac{\exp(\pi_{it} - v_{i(t-1)})}{\exp(\sum_k (\pi_{kt} - v_{k(t-1)})}$, and the exponential ratio between prediction error for the chosen action the maximum prediction error in the same trial, $\frac{\exp(\pi_{it} - v_{i(t-1)})}{\exp(\pi_t^{\max} - v_{(t-1)}^{\max})}$. Regardless of how the relative prediction errors are calculated for belief-based learning, the results were robustly similar.

In the rest of the paper, we adopted orthogonalized prediction errors for chosen actions against the maximal fictive error in the same trial as relative prediction errors for belief-based learning, unless otherwise specified.

2.4 Behavioral Results

Aggregate Behavioral Results

To measure the effectiveness of our behavioral protocol, we compared both aggregate choices between (1) our neuroimaging subjects, (2) our behavioral subjects, and (3) Rapoport and Amaldoss's (2000) original experiment (Table 1). We found no evidence that the neuroimaging protocols substantially affected players' behavior.

Individual Level Behavioral Estimates

To address which of the three learning models most closely capture subjects' behavior, we fit each model to behavior separately and compare the goodness of fit. Table 2 shows median individual-level estimates with associated first and third quartiles. Notably, parameter δ is significantly higher in median or mean individual level estimates than the pooled estimate reported in Rapport (2000), reflecting downward bias induced through heterogeneity (Wilcox, 2006).

The results of behavioral estimation suggest a couple of important features. Firstly, consistent with previous empirical studies, EWA model on average outperforms both reinforcement and belief-based in fitting the behavioral data in both in-sample tests using the Bayesian information criterion penalizing for number of free parameters (Figure 7A), as well as out-of-sample tests (Figure 7B). Figure 8 uses simulated play

based on best-fitting estimates to show a subject who clearly exhibits both aspects of reinforcement and belief-based learning. The notable aspects of this plot is that reinforcement model misses the increased probability of investing 4 in rounds 20-40 (Figure 8C), but this is captured by the belief-based learning model (Figure 8D). On the other hand, the belief-based learning model overestimates the probability that strategy 4

			Empirical Distributions		
Role	Investment	Equilibrium Prediction	Matrix Form	Behavioral Session	Neuroimaging Session
Strong	0	0%	1%	0%	1%
	1	20%	17%	14%	18%
	2	0%	5%	6%	10%
	3	20%	9%	13%	11%
	4	0%	13%	25%	16%
	5	60%	55%	43%	45%
Iak	0	60%	55%	49%	49%
	1	0%	3%	3%	4%
	2	20%	6%	10%	7%
	3	0%	14%	10%	14%
	4	20%	22%	28%	27%

Table 1: Comparison of Nash equilibrium predictions and empirical distributions from (1) Rapoport and Amaldoss (2000), (2) our behavioral experiment, and (3) our neuroimaging experiment. Empirical distribution is proportion of all players' choices over all rounds.

Model	δ	φ	λ
reinforcement learning	0*	.94 (.86, .96)	.04 (.02, .07)
belief-based learning	1*	.95 (.83, .98)	.60 (.23, 2.11)
EWA	.46 (.29, .69)	.71 (.53, .81)	.51 (.32, .70)

Table 2: Median individual level estimates. Parentheses contain first and third quartile of empirical distribution.

will be played in rounds 40-60 (Figure 8D), whereas this is not the case with the reinforcement model (Figure 8C). These are precisely the features of the two models that EWA combines (Figure 8B).

Second, individual variability in learning was further quantified using the estimated EWA parameter $\hat{\partial}_i$, which captures the relative weights placed on RL and BB learning (Figure 7C). Crucially for interpretation of neuroimaging results, we found that behavior in the game was driven by participants engaging in *both* RL and BB learning at

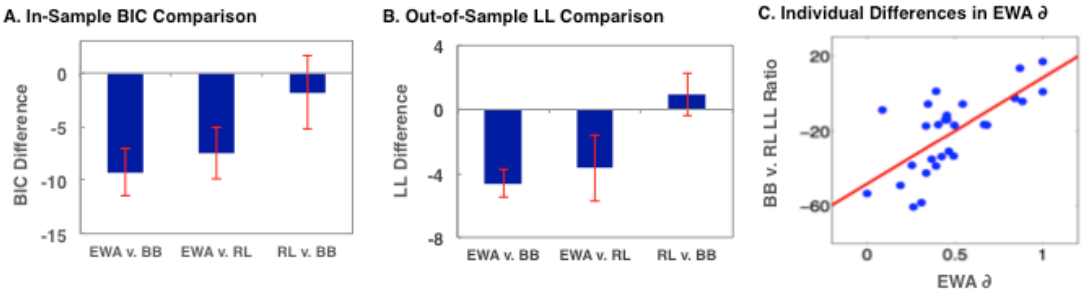

Figure 7: Behavioral results. (A) BIC comparison of in-sample model fit showed that EWA fits behavioral choices significantly better than RL and BB models ($p \leq 0.01$, paired t-test, two-tailed). (B) EWA also had superior out-of-sample predictive power ($p \leq 0.01$, paired t-test, two-tailed). Error bars are SEM. (C) Individual variation in the relative weights placed on RL and BB learning can be captured using individualized EWA parameter $\hat{\partial}_i$. That is, as ∂ increases, behavioral fit of BB learning improves relative to that of the RL (Pearson $\rho = .70, p < 0.01$, two-tailed).

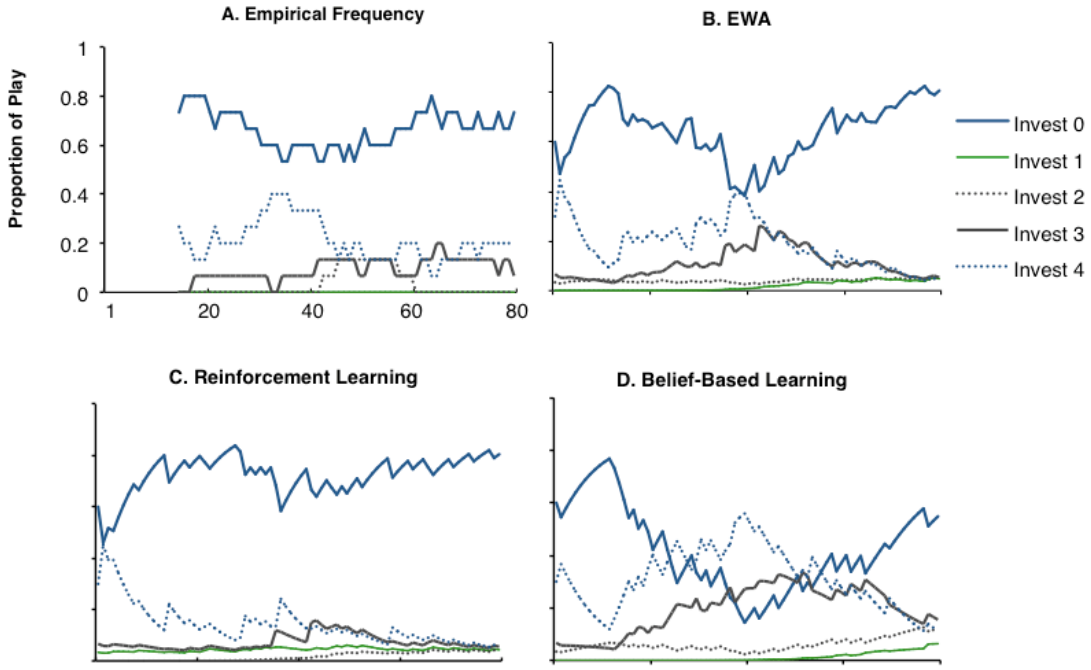

Figure 8: Comparison of empirical frequency and model fits for a subject predicted to exhibit both reinforcement learning and belief-based learning ($\delta = 0.47$). (A) Empirical frequency of play calculated using a 15 round window, (B-D) Estimated probability of play with reinforcement learning, belief-based learning, and EWA models.

the individual level, rather than a mixture of distinct segments of pure RL and BB learners. That is, individual $\hat{\partial}_i$ values were distributed along the unit interval rather than clustered at the boundaries as would be expected with distinct segments of pure RL and BB players (Figure 7C, Table 2). Furthermore, this variation allowed us to use $\hat{\partial}_i$ as a between-subject measure in subsequent neuroimaging analysis.

Correlation of Prediction Error Regressors

After obtaining the best-fitting estimates for each model, we convert the models into corresponding TD forms (see 2.3 for details) to compute corresponding model generated prediction error signals for each subject in each trial. Table 3 presents the correlation coefficients between these prediction error signals derived from

reinforcement, belief-based, and EWA. As shown in Table 3, the correlation between the prediction errors of reinforcement and belief-based learning is quite low (mean Pearson $\rho \approx 0.28$). This is especially important when these prediction error signals are entered as regressors into our neuroimaging data analysis. The statistical separation between the model-generated learning signals indicates the potential to disentangle the unique contributions of the different types of learning signals. The correlation of reinforcement and belief-based learning with EWA is also as expected; given the two are nested models within EWA.

	RL	BB	EWA
RL	-	(.16)	(.10)
BB	.28	-	(.18)
EWA	.63	.40	-

Table 3: Correlation coefficient between the prediction errors from different learning models. Parentheses contain standard deviations for the correlation coefficients.

2.5 Neuroimaging Results

RL and BB Prediction Errors

To relate the learning algorithm to brain activity, we use the trial to trial prediction error signals generated by the learning models as regressors on BOLD signals acquired at the time of the outcome, using statistical parametric mapping (SPM) to identify brain regions in which neural activity was significantly correlated with the model's internal signals.

First we look for brain regions that respond to the reinforcement prediction errors. This is the standard TD reinforcement error signal found in many previous studies of

reward learning (McClure, Berns et al. 2003; O'Doherty, Dayan et al. 2004). We find that the reinforcement prediction error signal correlates with activity in the ventral striatum (putamen) bilaterally (Figure 9A), consistent with many previous findings implicating this area in TD error encoding.

Next we search for regions showing responses consistent with the prediction error signal from belief-based learning. Two versions of prediction error signals are considered. The first one is Cournot error, which is the one-period regret, defined as the difference between the highest possible payoff and the received payoff. The second one is the belief-based prediction error signal derived from the dynamic belief-based learning model introduced in part 2.2. We find that Cournot error appears to be correlated only with activity in the bilateral putamen (Figure 9B), whereas the prediction errors from the belief-based learning model significantly correlate with the activity of both the bilateral putamen and a large region that extends across the medial prefrontal cortex (mPFC), orbitofrontal cortex (OFC), and anterior cingulate cortex (ACC). (Figure 9C).

We then examined the individual differences in the degree of belief-based learning within our subject group, by comparing the difference in the estimated value of δ obtained from individual EWA estimations and correlating that with neural activity elicited only by the belief-based prediction error signal. A significant between-subject correlation is found in the ACC and part of the mPFC. This result suggests that among subjects who assign higher weights to belief-based learning, the prediction error signals derived from their belief-based model correlated better with their neural activities in the ACC and the mPFC. The scatter plot in the lower right panel of Figure 9C shows this is indeed the case.

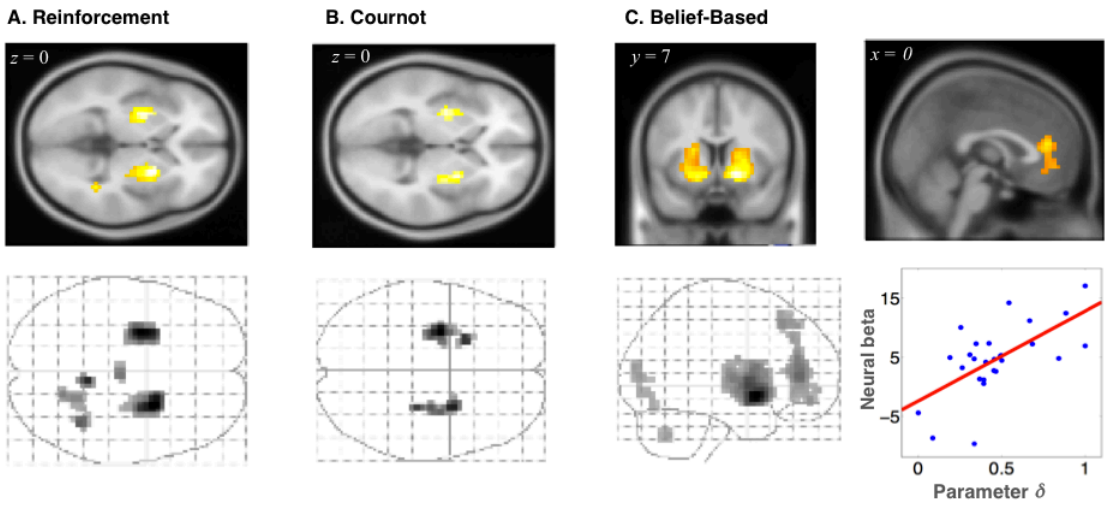

Figure 9: (A) Activation of the bilateral putamen in response to prediction error signals in reinforcement learning model ($p < 0.001$ uncorrected, cluster size > 10). (B) Activation of the bilateral putamen in response to prediction error signals in Cournot model ($p < 0.001$ uncorrected, cluster size > 10). (C) Significant activation of the medial prefrontal cortex and the bilateral putamen to belief-based prediction error ($p < 0.001$ uncorrected, cluster size > 10). Brain behavior correlation of the δ parameter estimated behaviorally against the fMRI estimates at ACC (Pearson = 0.66, $p < 0.001$).

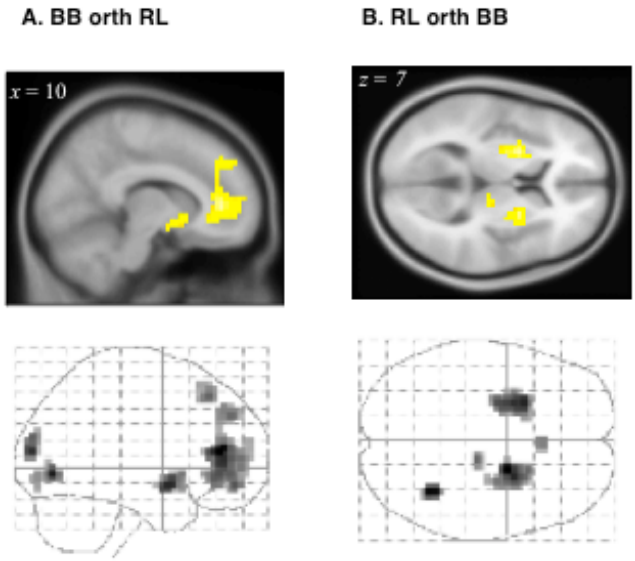

Figure 10: Robustness check for orthogonalization between RL and BB models. **(A)** BB relative prediction errors after orthogonalization against RL prediction errors. ($p<0.001$, uncorrected , cluster size $k > 10$ voxels). **(B)** RL prediction errors after orthogonalization against BB relative prediction errors. ($P<0.0001$, uncorrected, cluster size $k > 10$ voxels).

We tested the robustness of the result in the following ways. Firstly, we simultaneously included the reinforcement prediction error and the orthogonalized belief-based prediction error in the model with both regressors time-locked to the moment when outcome is revealed. We verified that the result is robust against the variation of regression models thanks to the low correlation between the prediction error signals from respective learning models (Figure 10).

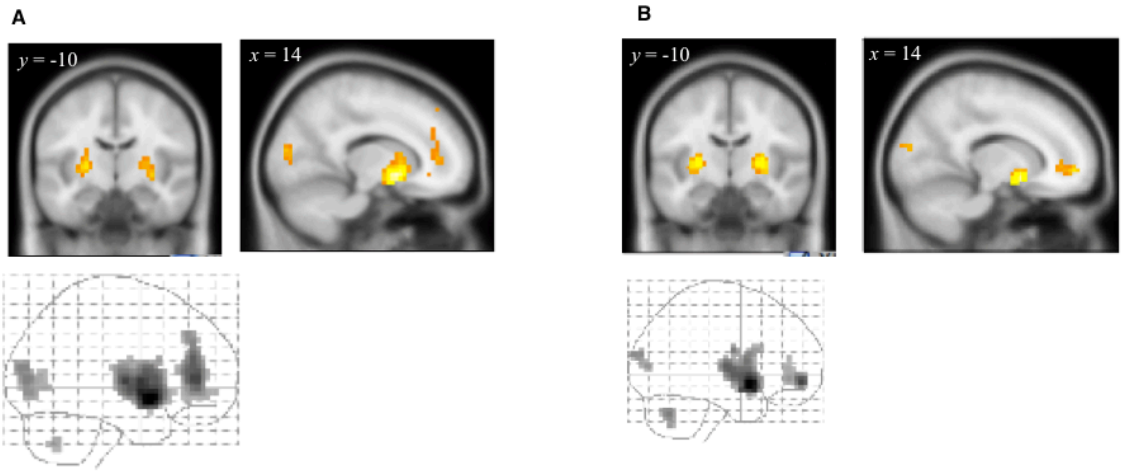

Figure 11: Robustness check for the notion of relative belief-based prediction errors. (A) The orthogonalized prediction errors from chosen action against average prediction errors for all available actions in the same trial. (P<0.001, uncorrected, k>10) (B) Exponential ratio between the prediction error for chosen action and the maximal prediction error among all prediction errors for available actions in the same trial. (P<0.005, uncorrected, k>10)

Secondly, we verified that the result for the belief based prediction error does not depends on the specific notion of relative prediction error we used. We therefore tested the following two alternative definitions of relative prediction errors: the orthogonalized prediction errors from chosen action against average prediction errors for all available actions in the same trial as well as the exponential ratio between the prediction error for chosen action and the maximal prediction error among all prediction errors for available

actions in the same trial. As shown in Figure 11, same brain regions respond to those notions of relative belief-based prediction error signals.

Unitary vs. Dual System Test

Having verified the spatial segregation in brain regions underlying the RL and BB prediction errors, we next explored the presence of a unified hybrid prediction error. Being a hybrid model, such prediction errors are by construction correlated with both RL and BB prediction errors. As in the behavioral results, however, if the true model of the underlying neural activity was a single hybrid prediction error, the hybrid prediction error should fit better than either RL or BB alone. In particular, the putamen would appear to be an obvious candidate if the brain encoded a hybrid prediction error given its involvement in both RL and BB prediction errors. In order to test this conjecture, next we searched for brain regions in which neural activity was significantly correlated with the EWA learning signals. Figure 12 contains the brain regions where the neural signals covariate with the EWA prediction error signals, which fails to support the hypothesis for putamen as the "EWA region", as the same array of regions were implicated in the hybrid as was in BB and RL prediction errors.

To test this formally, we first searched for brain regions showing a significantly better fit to the joint RL and BB prediction errors. We orthogonalized both RL and BB learning prediction error signals against the hybrid prediction error signal and conducted a F-test with the two orthogonalized regressors. We found that RL and BB had additional explanatory power, controlling for EWA, in the bilateral putamen and rACC (Figure 13A). In contrast, in the reverse analysis, we did not find the orthogonalized hybrid

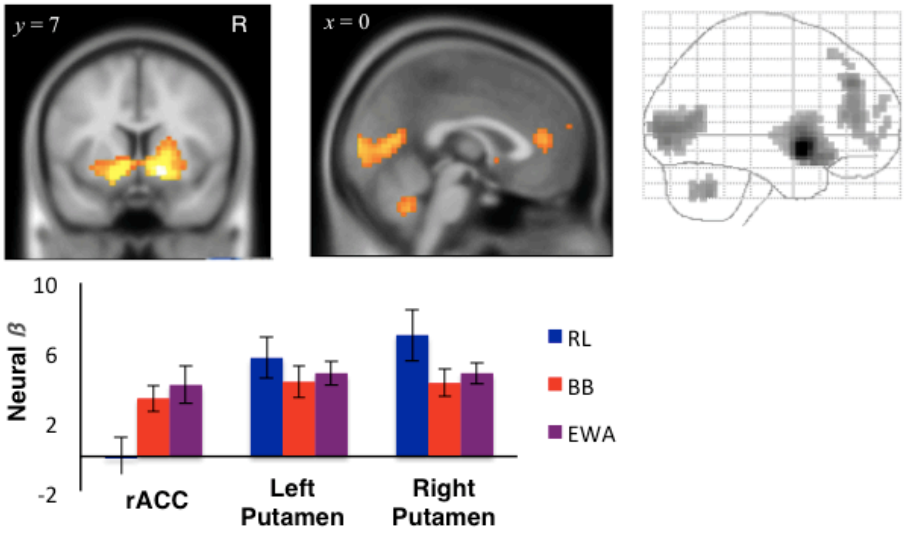

Figure 12: EWA prediction error signal. Up: Glass brain and section of bilateral putamen rACC activation in response to EWA prediction error ($p < 0.001$ uncorrected, cluster size $k \geq 10$). Down: neural betas with respect to reinforcement, belief and EWA prediction errors at three ROIs defined by 8-mm sphere at the peak of EWA activation clusters (right putamen: 14, 7, -7, left putamen: -14, 7, -10, rACC: 10, 56, -4).

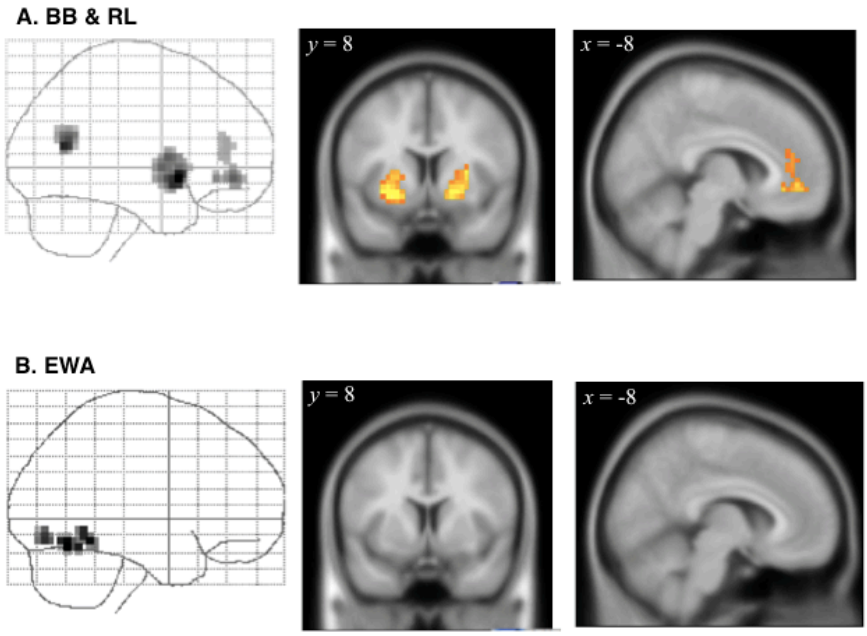

Figure 13: (A) Orthogonalized reinforcement and belief learning prediction errors remain significantly correlated with putamen and rACC activity ($p < 0.001$ uncorrected, $k \geq 10$), whereas (B) orthogonalized EWA prediction errors showed no significant correlation in these regions ($p < 0.001$ uncorrected, $k \geq 10$).

prediction error correlated with activity in any of the regions traditionally associated with decision-making and learning. That is, the hybrid prediction error did not have extra explanatory power, controlling for RL and BB, except in a small area of the visual cortex (Figure 13B). This is the case even at a liberal threshold of $p < 0.01$, uncorrected. Taken together, the above evidence rejects the hypothesis of a single-process hybrid prediction error in the brain.

We checked the validity of the above test in a number of ways. First we addressed the concern that after orthogonalizing against both RL and BB prediction errors, the hybrid prediction errors would be pure noisy signals, which would be particularly true if the hybrid prediction error was a simple linear combination of RL and BB prediction errors. We examined the correlation between the orthogonalized and the unorthogonalized EWA prediction errors for each subject playing each role. The average correlation coefficient was around 0.40 ($p < 0.05$), suggesting that the orthogonalized hybrid EWA prediction error signals preserved a substantial portion of the unique share of variance after taking away the shares from RL and BB learning signals.

Next we verified that the non-activation of the orthogonalized EWA prediction errors is not due to the over-penalization of EWA by orthogonalizing its prediction errors twice against RL and BB. We applied the similar orthogonalization strategy to RL and BB prediction errors respectively (i.e. orthogonalized RL prediction error against both EWA and BB prediction errors, and orthogonalized BB against both RL and EWA prediction errors). Their activation regions are overlaid with the unorthogonalized EWA activation regions in Figure 14.

Possible Mediating System for the Multiple Prediction errors

Further analysis is conducted on the between-subject variability in the degree to which the dual-process is mediated. Because subjects who have interior values of δ will require cognitive mediation of the respective learning signals, we use $\delta(1-\delta)$ to approximate the level of such cognitive mediation between the dual-process of learning for each subject. Brain regions invoked in mediating the dual-process will correlate more strongly with the orthogonalized EWA prediction error signal for subjects employ such cognitive mediation when compared with subjects that do not. We conduct between subjects analysis to look for the regions in which neural betas for the orthogonalized EWA prediction errors correlates with $\delta(1-\delta)$. Mediating signals are found to significantly covary with the orthogonalized EWA prediction errors across subjects in the dorsolateral prefrontal cortex (DLPFC) and inferior parietal lobule (IPL) (Figure 15), revealing these regions possible importance in mediating the balance between reinforcement and belief-based learning.

However, such a result is very limited, give that it is a between subject analysis. Future study will be done to test the explore this part more formally including PPI on a trial by trial basis within each subject.

Expected Reward Regions

As shown in Figure 16, we found activity in ventromedial prefrontal cortex, extending to rACC and medial orbitofrontal cortex, to be correlated with the relative

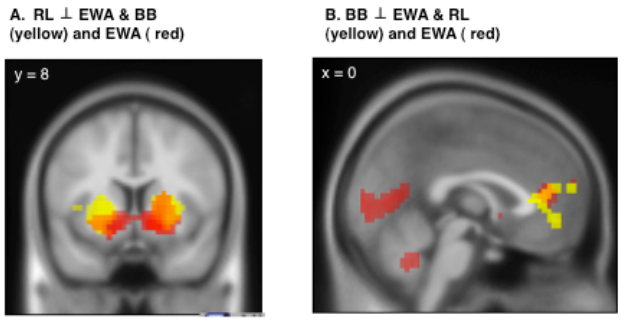

Figure 14: Region significantly correlated with RL prediction errors after orthogonalizing it against EWA prediction error and BB relative prediction errors (yellow) overlays with activation region for EWA prediction errors (red) ($p < 0.001$, uncorrected, cluster size $k > 10$ voxels). **(B)** Region significantly correlated with BB relative prediction errors after orthogonalizing it against EWA and RL prediction errors (yellow) overlays with activation region for EWA prediction errors (red) ($p < 0.001$, uncorrected, cluster size $k > 10$ voxels). Notably, activation in the putamen is no longer correlated with the BB relative prediction errors, controlling for EWA and RL. The only activation region for the double orthogonalized BB relative prediction errors is rACC.

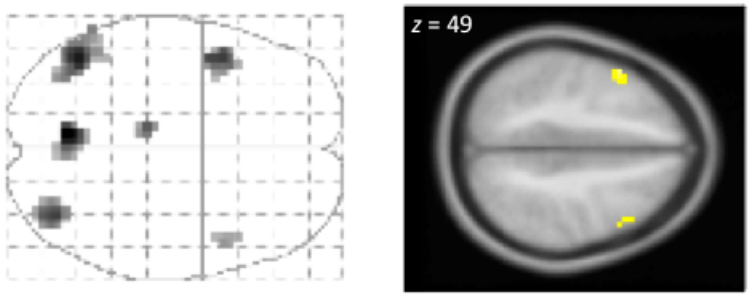

Figure 15: The degree to which a subject mediates reinforcement and belief-based prediction error signals is measured by the estimated value of $\delta(1-\delta)$. Activities in the dorsolateral prefrontal cortex (DLPFC) and inferior parietal lobule (IPL) correlates better with the orthogonalized EWA among subjects with more intensive cognitive mediation between the two separate learning processes ($p < 0.001$ uncorrected, cluster size > 10)

expected reward value of the chosen action. The relative expected reward is defined as the probability generated from the EWA model for the chosen action on a given trial. We used this notion in order to remove the possible time trend in the absolute expected reward values. This result is consistent with existing evidence on the role of orbital and adjacent medial prefrontal cortex in encoding predictions of future reward (Daw, 2006; O'Doherty, 2004). It further supports the hypothesis that strategic learning can be characterized as a TD process.

Other Forms of BB Prediction Errors

In principle, BB model does not rule out the possibility that the brain encodes prediction error signals beyond those of chosen strategies at the feedback event. This is particularly interesting because in BB learning the expected values of unchosen actions do not decay at a rate constant across all unchosen actions as in the standard RL, but rather vary in a relatively complicated way depending on the change of player's belief about opponents' future strategy as well as the payoff of each action profile. Figure 17 shows the activation regions correspond to different notions of relative prediction errors for BB learning beyond those of chosen actions. This evidence, however, is limited because our design was not optimized to find prediction errors other than those associated with the chosen actions.

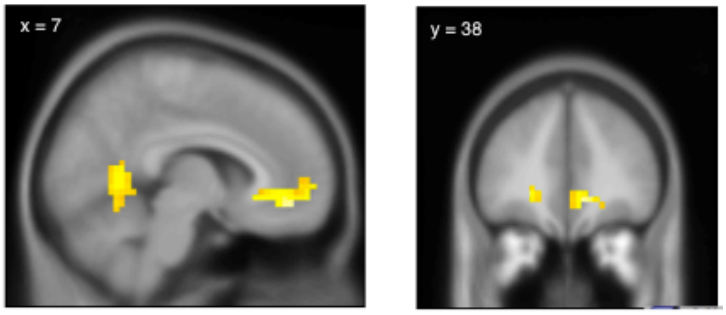

Figure 16: Activity in ventromedial prefrontal cortex, extending to rACC and medial orbitofrontal cortex, is correlated with respect to relative expected reward value of the chosen action calculated under the EWA model ($p < 0.005$ uncorrected, cluster size > 10).

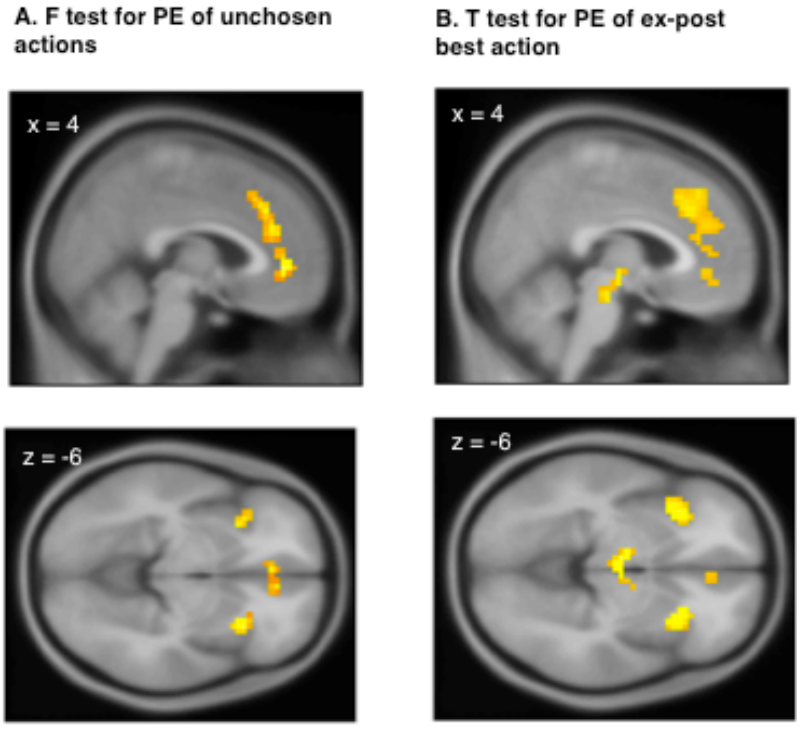

Figure 17: (A) F-test for regions significantly correlated with BB relative prediction errors of unchosen actions for weak players ($p < 0.005$ uncorrected, cluster size > 10). (B) Regions significantly correlated with BB prediction error of the best ex-post action for weak players ($p < 0.001$ uncorrected, cluster size > 10)

Mentalization

Beyond rACC, we found no evidence of involvement of those regions commonly implicated in mentalization, such as temporoparietal junction (TPJ) (Young, Cushman et al. 2007) and superior temporal sulcus (STS)(Amodio and Frith 2006). This is consistent with the view that there exist different processes that meet different computational demands underlying mentalization (Jenkins and Mitchell 2010). That is, rACC activity in our population game may reflect first-order belief formation that is critical for belief-based learning, whereas regions such as TPJ and STS may only come online in higher-order mentalization and belief inference. This is consistent with findings from a fixed-pair matching-penny game that taps higher-order belief inference (Hampton, Bossaerts et al. 2008), although another study that uses a spatial game did not find such regions (Yoshida, Dziobek et al. 2010).

2.6 Discussion

In order to test for the strategic learning process among a set of candidate models, the conventional method is to compare the in-sample explanatory power of different competing models that provide the best account of behavior. This runs into problems when two or more competing models provide qualitatively similar predictions about behavior. Using simulations, Salmon (2001) found this to be a potentially serious problem in traditional experimental studies on strategic learning. Here by taking the advantage of recent breakthroughs in understanding of the neural underpinnings of decision-making, we provide a new dimension of data and a biologically plausible criterion for understanding the strategic learning process. By disentangling the unique

contributions of various candidate learning models at neural level, we aim to (1) distinguish the underlying strategic learning process more accurately and to (2) provide insights on improving the existing models on strategic learning.

In the present study, we explored the neural mechanisms underlying strategic learning by investigating the following two questions: (1) does there exist neural mechanisms that encode learning signals in the manner predicted by economic models of strategic learning? (2) If so, does there exist a single unified neural system that drives strategic learning, or is behavior driven by inputs from possibly separable systems? Unlike previous studies of strategic learning, we designed our study to minimize collinearity in the outputs associated with the different models under consideration. We found that human participants learn though separate reinforcement and belief-based processes. These distinct signals were represented in both overlapping and distinct brain regions. Our study is therefore the first that provides evidence for model-based belief learning in a competitive game. Furthermore, by using non-nested model hypothesis testing, we was able to reject the hypothesis that learning behavior is driven by a unified model such as EWA.

More specifically, we found that several learning signals—reinforcement, Cournot, and belief-based prediction errors—all appear to be represented in the putamen. The putamen is a region known to be involved in decision-making and learning (Yamada, Matsumoto et al. 2004; Daw, O'Doherty et al. 2006). A number of previous neuroimaging studies on non-strategic reward learning found that activity in the putamen to be correlated with reward prediction error (McClure, Berns et al. 2003; O'Doherty, Dayan et al. 2003; O'Doherty, Buchanan et al. 2006; Rangel, Camerer et al. 2008). More

recently, Lohrenz et al. (2007) studied the neural correlates of foregone payoff and found large activations to foregone payoffs in both ventral caudate and ventral putamen in a stock market investment task, suggesting that striatal regions including caudate and putamen are sensitive to foregone outcomes generally.

Second, we found regions that dissociate between reinforcement and belief-based learning. In particular, we found that belief-learning signals were encoded in the anterior cingulate (ACC) and the medial prefrontal cortices (mPFC). This finding differs with the previous studies on learning with no strategic or social motivations, where activations in ACC and mPFC are typically absent. In contrast to striatal activity, the ACC, particular the rostral portions found in our study, and mPFC are more commonly associated with higher order executive functions such as control (Hampton, Bossaerts et al. 2008; Lee 2008; Coricelli and Nagel 2009) and mentalizing (Hampton et al., 2008; Coricelli et al., 2009), respectively. Given extensive reciprocal anatomical connections between the striatal (which includes putamen) and more prefrontal regions, it is difficult to hypothesize the degree to which either or both regions are necessary for representation of foregone payoffs/beliefs. Disentangling this therefore requires techniques capable of assessing causality, such as the lesion method.

Importantly, we reject the hypothesis of a hybrid EWA process at the neural level, even though it outperforms reinforcement and belief-based learning models behaviorally. This supports the hypothesis that behavior emerges as a result of multiple and possible competing learning signals. These signals are in turn encoded in several unique and overlapping regions. This raises the interesting question whether there exists a third system that mediates the different learning signal inputs. This has implications given the

long-running debate on whether economic behavior is a result of the output of a unitary system or that of a competitive process between different brain systems (McClure, Laibson et al. 2004; Kable and Glimcher 2007; McClure, Ericson et al. 2007). Although our task was not optimized for this purpose, we provide some preliminary evidence that suggest the involvement of the DLPFC and IPL in such a mediating process. Specifically, individuals that exhibited both reinforcement and belief-based learning behaviorally showed greater activation in the DLPFC and IPL than those who exhibited primarily reinforcement or belief-based learning. This is consistent with previous studies that implicate DLPFC and IPL in cognitive regulation (MacDonald, Cohen et al. 2000; Badre and Wagner 2004).

Our results have important implications for both economics and neuroscience. For neuroscience, we provide a potentially useful paradigm to study social and learning deficits in a variety of mental and neurological illnesses. For economics, we show how a combination of traditional laboratory experiments and newly available methods from neuroscience can provide novel data about latent decision-making processes. This will potentially allow us to improve predictive power of econometric models of learning dynamics in real-world situations.

CHAPTER 3

NEURAL CAUSALITY FOR STRATEGIC LEARNING: A LESION STUDY

3.1. Introduction

The neuroimaging method we used in the previous chapter is limited by its observational and correlative nature. It only provides inferences about the association of the brain regions with the learning signals, and hence it is unable to conclusively demonstrate whether the identified neural system is necessary for strategic learning. In particular, our fMRI study shows that both putamen and anterior cingulate cortex are involved in belief-based learning. Given the extensive reciprocal anatomical connections between these regions, however, it is difficult to tell the degree to which either or both regions are necessary for representation of belief learning signals based on the neuroimaging result. More importantly, the fMRI study showed that strategic learning involves dissociable learning signals, which independently activate both distinct and overlapped brain areas known to process reward learning signals and social belief inferences. Yet it does not prove the brain areas activate by different learning signals are necessary for the putatively isolated learning processes. There remains a critical gap in the evidence relating strategic learning, reward learning and the brain.

Here we adopt the lesion method to disentangle the causality of different brain regions for strategic learning by examining whether the orbitofrontal cortex, basal ganglia, and dorsolateral prefrontal cortex are necessary for strategic learning. In particular, we compared the behavior of patients with lesion to orbitofrontal cortex (OFC,

n = 7), basal ganglia (BG, n = 6), and dorsolateral prefrontal cortex (DLPFC, n = 9) across strategic learning treatment and a control treatment of reward learning.

Hypotheses

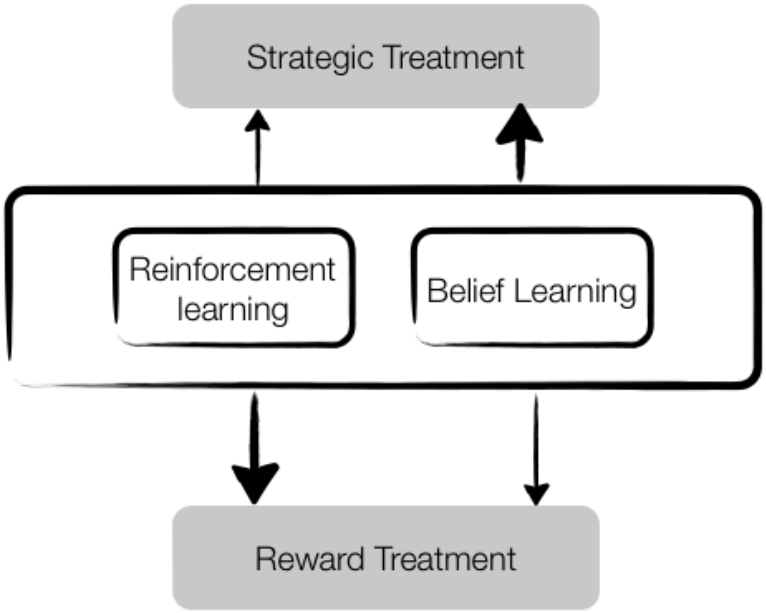

Figure 18: Hypothesis for strategic and reward learning. Both involve reinforcement and belief based learning, with different weights.

Based on both the literature we reviewed and our preliminary neuroimaging analyses, We hypothesized that both strategic and reward learning can be decomposed to reinforcement and belief-based (fictitious play) learning, with higher weight on belief-based learning relative to reinforcement learning in the strategic treatment than in reward learning treatment (Figure 18). We further hypothesized that there exist double dissociation between different brain regions and different types of learning. In particular, basal ganglia (which includes putamen) are necessary for reinforcement learning and the OFC necessary for belief-based learning. Furthermore, the DLPFC are necessary for

mediating reinforcement and belief-based learning processes in both strategic and reward treatments. Based on these hypotheses, we predicted that patients with basal ganglia lesions will be impaired in reward learning treatment, patients with the OFC lesions will show deficits in strategic learning treatment, and patients with the DLPFC lesions will be impaired in both reward and strategic learning.

3.2. Background

Lesion Study

Lesion studies seek to draw necessity inferences by demonstrating the inactivation of particular brain regions disrupt a certain cognitive process. It has been proven to be a powerful method especially when combined with functional neuroimaging.

The first group of lesion patients we consider is the orbitofrontal cortex and ventromedial prefrontal cortex (VMPC) patients. The region projects to basal forebrain and brainstem area, and neurons within the OFC and VMPC have been shown to encode the emotional, social value of sensory stimuli (Stone, Baron-Cohen et al. 1998; Tekin and Cummings 2002; Rudebeck and Murray 2008). Patients with damage to orbitofrontal cortex and with ventromedial damage, typically have severe deficits in social and decision-making functioning. For example, in the famous case of Phineas Gage, he survived the accident of an iron rod blast that damaged the bilateral OFC area of his brain. Surprisingly, his intelligence, memory and other cognitive abilities were intact, but social and decision-making abilities were found to be impaired (Ongür and Price 2000). Over the past decade this interesting pattern of impairment has been confirmed from other OFC lesion patient studies and dysfunction of the OFC has been further

characterized by personality change including lack of interpersonal sensitivity and empathy, and responding inappropriately to social clues (Bechara, Tranel et al. 1997; Stone, Baron-Cohen et al. 1998; Camille, Coricelli et al. 2004; Maia and McClelland 2004; Koenigs, Young et al. 2007; Wallis 2007; Chamberlain, Menzies et al. 2008; Krajbich, Adolphs et al. 2009).

The second group of lesion patients we consider is basal ganglia (which includes putamen) patients. There have been robust evidences suggesting that basal ganglia are crucial in stimulus-response learning (Bellebaum, Koch et al. 2008; Yehene, Meiran et al. 2008; Baier, Karnath et al. 2010; Maia and Frank 2011). For example, according to the actor-critic model, which have used to account for many behavioral and neural findings in stimulus response learning for both human and nonhuman primates, the basal ganglia is critical for action selection by implementing two learning modules: (1) the critic, which learns state values, and (2) the actor, which learns stimulus-response associations (O'Doherty, Dayan et al. 2004). Some evidences further suggested that the basal ganglia likely select actions according to the probabilities learned from reward reinforcement among candidate actions that are initially generated by the cortex (Frank 2005; Frank, Woroch et al. 2005). However, there is not much well established experimental probe for decision making for patients with basal ganglia lesions, mostly due to the scarcity of the patients. Studies for patients with Parkinson's disease, which exhibits reduced striatal dopamine resulting from dopaminergic cell death, mainly adopted experimental paradigms of reward learning such as probability learning tasks and showed behavioral abnormalities consistent with the neural model predictions (Frank, Seeberger et al. 2004; Frank, Samanta et al. 2007) .

The third group we considered is the patients with dorsolateral prefrontal lesions. In the previous chapter, we have some preliminary result suggesting that this is the region mediating the two learning processes. It is widely acknowledged that the dorsolateral prefrontal cortex mainly involves in executive functions, including cognitive control, emotion regulation, and working memory (Tekin and Cummings 2002; Mansouri, Buckley et al. 2007; MacDonald, Cohen et al. 2000; Miller 2000). Patients with DLPFC lesion have been robustly shown through clinical studies impaired performances in the Wisconsin Card Sorting Test, which requires regulating or shifting actions, and reversal learning (Milner 1963; MacDonald, Cohen et al. 2000; Nyhus and Barceló 2009). Recent neuroeconomic study (Hare, Camerer et al. 2009) also suggested that the DLPFC plays a critical role in the deployment of self-control in everyday decisions, such as choices over healthy and unhealthy snacks, through mediating multiple independent value signals. Furthermore, our previous imaging study has also suggested DLPFC as a candidate region for mediating RL and BB prediction error signals.

Computational method for lesion study

Traditionally, the lesion study was based on direct comparison of the behavior across different lesion groups (or patients and healthy controls). Recently computational methods have adopted in lesion studies, which allows us to explore the differences at deeper level by reasoning from behavior to its mechanistic causes, and also provides a way to test the computational modeling itself (Maia and Frank 2011). For example, one of the approaches involves fitting the behavioral data of both control and lesion patient groups separately to a computational model, which made explicit assumptions on the

internal process underlying the cognitive task, and then comparing the best fitting parameters and correlations between these parameters and disease severity (Figure 19).

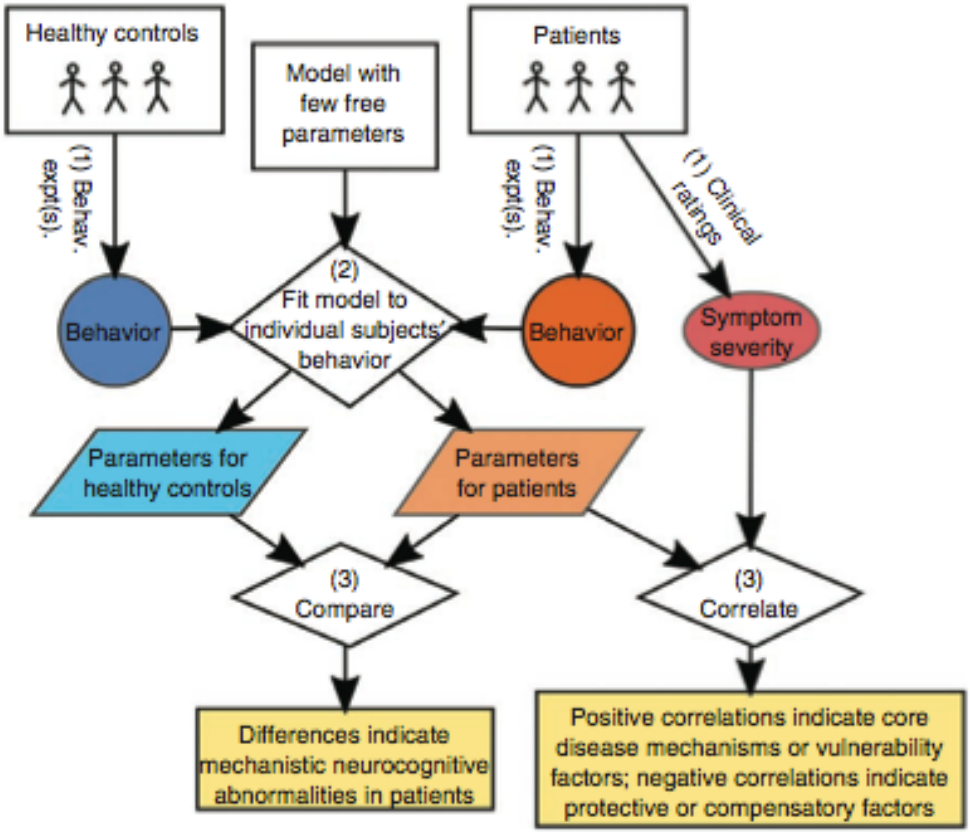

Figure 19: The "quantitative abductive" approach for computational method of lesion study. Adapted from Maia and Frank (2011).

3.3. Method

Subjects

We ran samples of 42 participants, including patients with orbitofrontal lesion (n = 7), basal ganglia lesion (n = 6), dorsolateral prefrontal lesion (n = 9), and healthy subjects (n = 10). All lesion patients are recruited from Bon Knight's Lab at University of California, Berkeley. The patients have stable, chronic lesions and are highly

functioning (Figure 20, individual lesion area are included in 3.6). Control subjects were recruited from craigslist and Berkeley recruiting center. They were matched to patients by age and education and screened according to health and neurological condition, usage of prescription drug and other medication. All subjects' demographical information as well as their memory performance, IQs and performance in Wisconsin Card Sort Test is listed in Table 4.

Procedure

All participants attended one strategic learning treatment and a reward learning treatments, with the order of treatments and roles in the game counter balanced. In the strategic learning treatment, participants played the Patent Race game, which was exactly the same as described in the previous chapter. All participants were randomly and anonymously matched with the pooled opponents (same pooled opponents as in the previous chapter) throughout the 80 rounds.

In the reward learning treatment, the participants were instructed that they were playing an 80 round card game against the computer. The rule and payoff structural is the same as in the strategic learning treatment. In each round the computer draws a card randomly from a deck containing either from Joker (equal to 0) to four or from Joker to five depending on the role of the player (Figure 21). Without knowing which card the computer has drawn, the player will pick a card from his own deck and he will win if his card is larger than the computer's card. Moreover, in each round the computer draws the card following the same probability as the empirical frequency of opponent's choices in

Group	Age	Gender # of Male	Education	Bower Memory	WCST Perseverative Errors
OFC (n=7)	47.57 (5.95)	3	15.14 (1)	9.43 (2.07)	6 (1.01)
BG (n=6)	64.00 (3.35)	4	14.33 (0.69)	4.67 (1.52)	4.67 (1.48)
DLPFC (n=9)	58.56 (2.33)	4	15.22 (0.98)	4.22 (1.39)	9.44 (1.74)
HC (n=10)	58.69 (3.24)	6	15.00 (0.59)	11.89 (1.93)	5.33 (1.03)

Table 4: demographical information of the subjects. Means (standard errors) of age, years of education, verbal memory score in bower memory test (out of 20) and perseverative errors in Wisconsin Card Sorting test (out of 48).

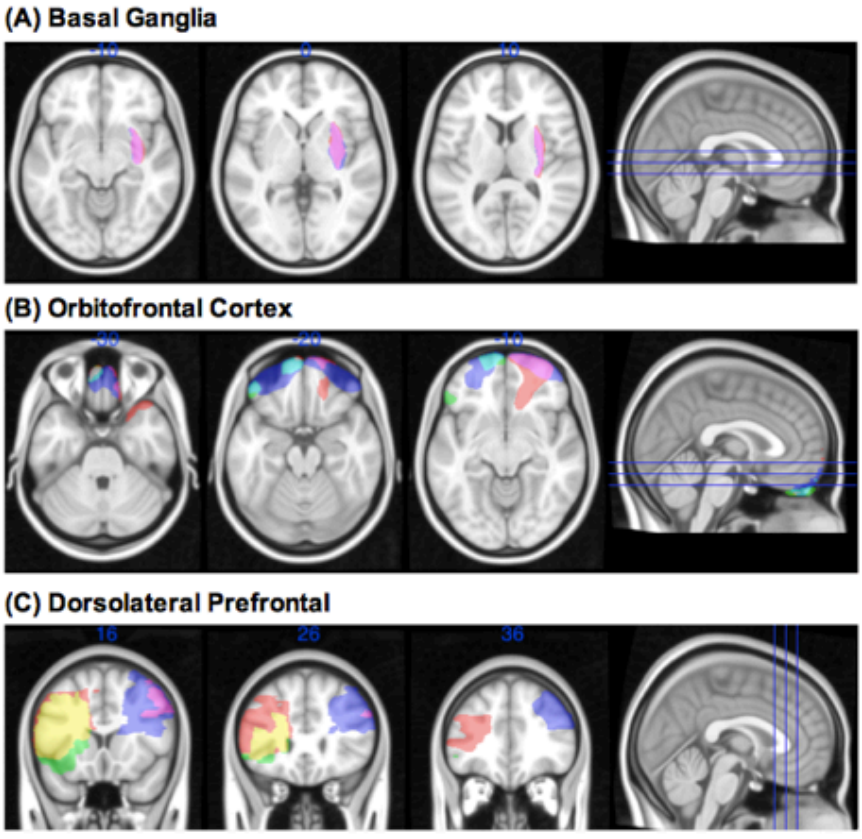

Figure 20: Patient MRIs. Structural MRI slices illustrating the lesion overlap across patient groups. Software reconstructions were performed using MRIcro (Rorden and Brett 2000).

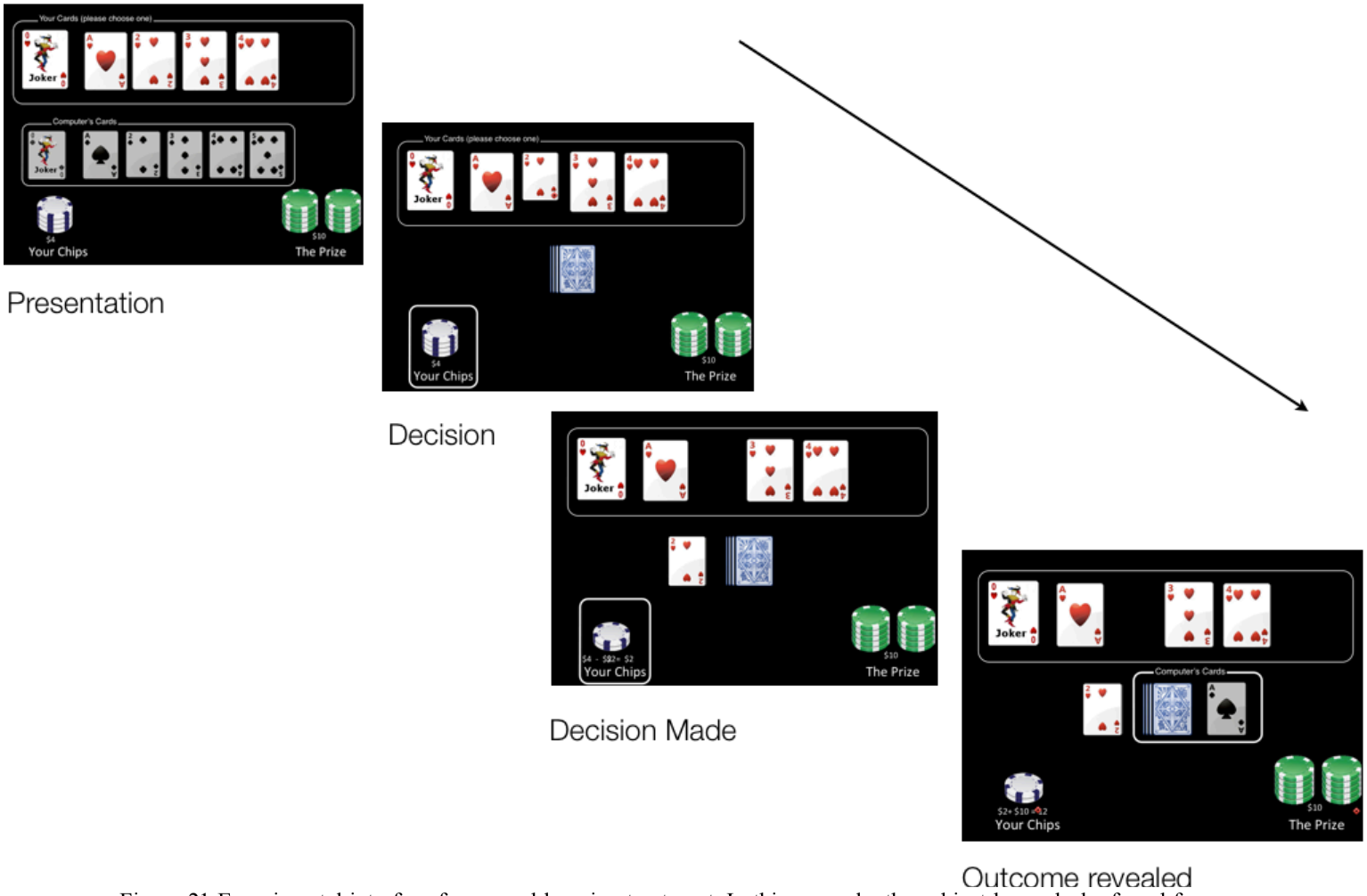

Figure 21:Experimental interface for reward learning treatment. In this example, the subject has a deck of card from Ace which is equal to 0) to four (i.e. weak role) and the computer has a deck of cards from Ace (which is equal to 0) to five. The payoff structural and the probabilities for drawing cards by computer are the same as in the strategic learning treatment.

the same round of the strategic treatment. Different interface was adopted only to avoid possible confusion between two treatments especially for those with possible memory impairments.

All participants received detailed instructions and practices, and were required to pass a quiz prior to the experiment to ensure sufficient understanding of rule of the game and the ability to predict payoffs given an arbitrary choice profile. Different from our previous fMRI study however, all participants were paid at a flat rate due to human subjects restrictions to lesion patients.

3.4. Results

Summary of aggregate empirical frequencies of choices are presented in Figure 5. A better way for visualization the choices is through the transition matrices (Figure 22), which show the joint empirical frequencies of choices at trial t+1 conditional on choices at trial t for each cohort in each role, computed by pooling the observations across trials and subjects within lesion group given under the particular treatment. These matrices show how players switch their choices from one trial to the next, and are generalizations of more traditional switch/stay measures. In particular, the diagonal elements indicate choices in which subjects stayed, whereas off-diagonals indicate switches.

Healthy Controls

It is clear through introspection that the behavior of the healthy comparison group in Figure 22 is quite similar between strategic and reward learning, except that more

	0	1	2	3	4	5
BG						
strategic	5.83%	12.50%	35.83%	27.08%	10.00%	8.75%
reward	0.42%	0.42%	0.42%	7.08%	48.33%	43.33%
DLPFC						
strategic	2.81%	3.75%	16.88%	32.50%	26.25%	17.81%
reward	15.00%	4.75%	7.25%	21.75%	27.00%	24.25%
HC						
strategic	7.19%	20.63%	16.56%	12.19%	8.75%	34.69%
reward	0.50%	23.75%	14.50%	36.25%	15.00%	10.00%
OFC						
strategic	8.75%	14.58%	20.42%	15.42%	12.50%	28.33%
reward	1.88%	25.94%	19.38%	19.38%	14.38%	19.06%

	0	1	2	3	4
BG					
strategic	15.00%	5.42%	12.92%	18.75%	47.92%
reward	27.50%	17.50%	8.75%	23.33%	22.92%
DLPFC					
strategic	46.25%	10.31%	14.69%	12.50%	16.25%
reward	5.94%	4.06%	15.31%	25.00%	49.69%
HC					
strategic	51.75%	1.25%	2.75%	12.75%	31.50%
reward	30.63%	5.00%	15.31%	25.00%	24.06%
OFC					
strategic	35.94%	7.19%	11.56%	23.75%	21.56%
reward	43.75%	2.08%	17.50%	11.25%	25.42%

Table 5: Pooled empirical frequencies of choices for each lesion group under each treatment. Up: strong role. Bottom: weak role.

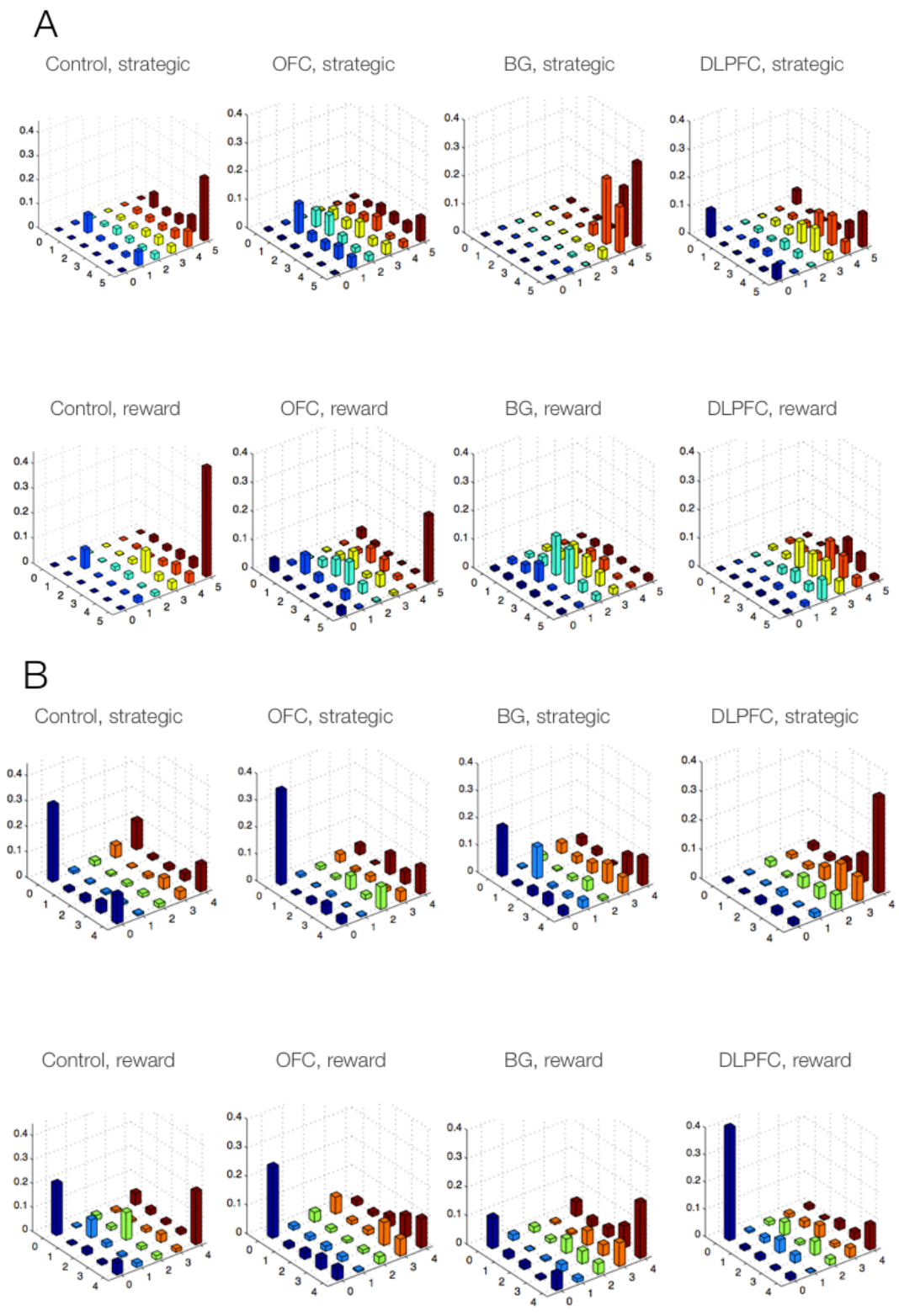

Figure 22: Transition matrices across lesion groups and treatments. On the X-axis of the matrix it is the empirical frequencies of the cohort's choice at trial t, and on the Y-axis, it is the empirical frequencies of the choice at trial t+1. A: Strong players. B: Weak Players.

weights are distributed on the diagonal of the transition matrices under reward learning treatments for both strong and weak roles. For example, strong players in the strategic learning treatment switched from investing 5 to 1 approximately 8% of all trials. In approximately 20% of all trials, Strong players stayed with investing 5. In contrast, the reward learning switching rate is much lower, consistent with the observation that reward learning, subserved primarily by reinforcement learning, is a slower process. This is apparent in that most of the mass of the transition matrix is located along the diagonal (indicating stay trials) at investment of 1, 3, and 5.

To understand why this is the case, we fit the EWA model with the experimental data from control subjects in reward and strategic learning respectively and conducted grid search over a large space of parameters. The result is visualized in Figure 25. The details of EWA model are introduced in the previous chapter. Consistent with our hypothesis, the estimation result shows that control subjects do present both reinforcement and belief-based learning in both strategic and reward treatments, yet with higher relative weight on belief-based learning under the strategic treatment (i.e. higher value of δ).

To further verify that the distinction in the choice patterns across two treatments can be explained by different portions of reinforcement and belief-based learning, we simulated choices under pure belief based learning and pure reinforcement learning respectively and visualized those with transition matrices. As shown in Figure 23, strategic learning is qualitatively more similar to belief based learning and reward learning to reinforcement learning simulations. More specifically, almost all the weights are distributed on the diagonal in the transition matrix for simulated reinforcement

learning, similar to the distribution of reward learning matrix. Whereas in the transition matrix for belief based learning the weights (except (5,5)) are more off diagonal, similar to the distribution of strategic learning.

Lesion Subjects

For lesion subjects, let's first take a look at the model-free measures of behavior, which can give us a better qualitative picture of the data if the underlying model assumptions are violated, and a sense of the qualitative deficits of the particular lesion groups. As shown in Figure 22, the primary difference between strategic and

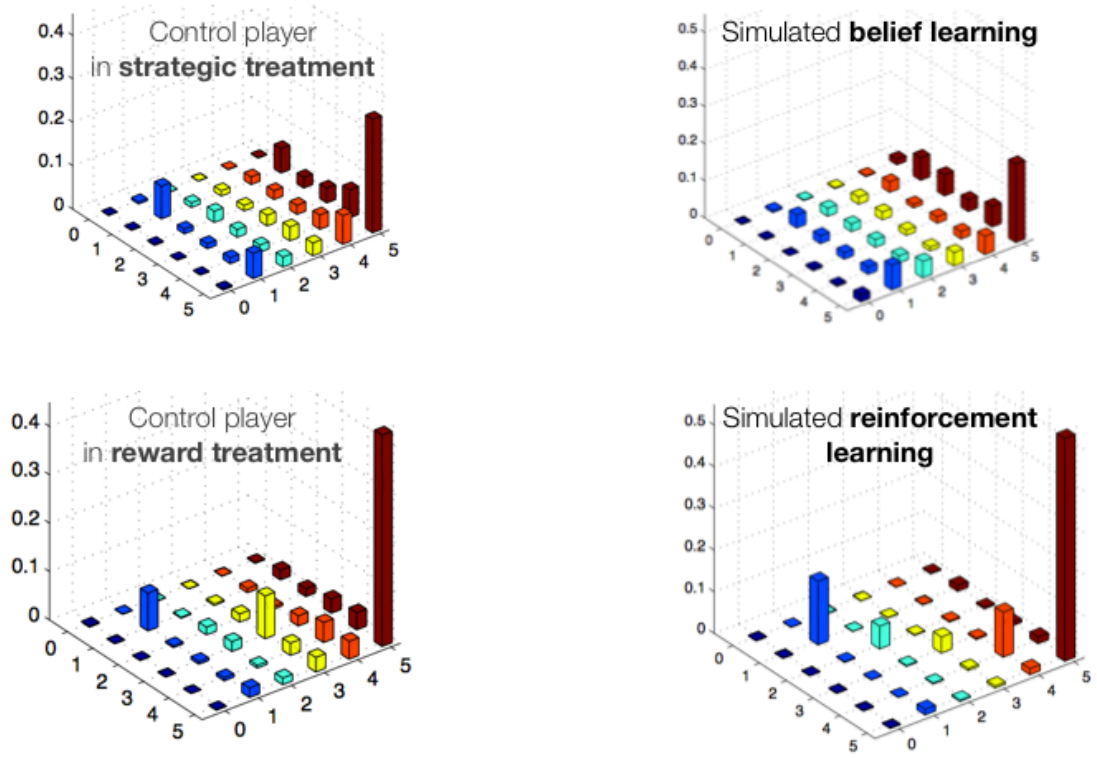

Figure 23: Comparison across behavioral and simulated data for control subjects playing strong roles. Top Left: Empirical frequency of transitions for behavioral data in strategic treatment. Top Right: transition matrix for simulated choices based on pure belief-based learning. Bottom Left: transition matrix for behavioral data in reward treatment. Bottom Right: transition matrix for simulated choices based on pure reinforcement learning.Model-free measure using transition matrices for Strong player.

reward learning in healthy controls is the switching behavior across rounds. Choices of the lesion groups however differ substantially across treatments, and support the hypothesis of distinct reinforcement and belief pathways. First, BG patients appear particularly impaired in reward learning, rarely choosing to invest 5 in the Strong role, instead often choosing 2 and 3. This can be compared to strategic learning, where they chose to invest 4 and 5 quite often. The OFC patients in contrast appear to show the reverse effort, performing seemingly randomly in strategic learning but quite well in reward learning. This raises the interesting possibility that the BG patients are able to use the belief based learning to substitute for the impairments in reward learning deficits.

Iteratively elimination of dominated strategies in strategic treatment

In order to further investigate the reasons for the distinct behavior patterns across lesion groups, it will be helpful to understand the game theoretic prediction about how a "perfectly rational" player would reason under the strategic treatment. The standard game theory arrives at a mixed strategy Nash equilibrium solution for the Patent Race game based on the assumption that both strong and weak players have perfect "theory of mind", so that they can introspectively imagine that they are in the shoes of the other player, and reason their way to Nash equilibrium in the following way: Consider first the choice of the strong player. The Strong player is guaranteed 10 by investing her entire endowment of 5. This is more than she can win by investing 0, which guarantees an earning of 5, the initial endowment. The Strong player, therefore, should never invest 0. Turning now to the Weak player, he now "knows" Strong will invest an amount strictly greater than 0. This leads Weak to never invest 1, as he will lose for sure (recall that both

players lose in ties) and earning 4 by 2 investing nothing is better than investing 1 and earning 3. Turning back to the Strong player, who knows Weak will not invest 1, will then never invest 2. This is because an investment of 2 is only optimal when the other party invests 1. This process continues until the Strong player removes all even strategies from consideration, Weak player all odd strategies, and both players randomize over the remaining strategies such that no player has a unilateral incentive to deviate. Such a process of reasoning is called iteratively elimination of (strongly) dominated strategies, which is one of the most basic solution concepts in games theory, and have been widely applied to experimental studies, such as Beauty Contest (Nagel 1995; Duffy and Nagel 1997), Centipede game (McKelvey and Palfrey 1992), and etc.

In reality, the first time players are exposed to a strategic game such as Patent Race or Beauty Contest Game, they are in general unable to conduct the full depth of iterative reasoning. Yet extensive experimental studies have shown that subjects are able to improve such reasoning when playing the game repeatedly. The question is that if lesion subjects are also able to improve strategic reasoning during learning.

The top panel in Figure 24 suggests different abilities of lesion participants to learn to eliminate iteratively dominated strategies in strategic learning, by comparing the empirical cumulative distributions for the portion of subjects playing the corresponding iteratively dominated strategies. For example, the first row of the strategic panel in Figure 8 suggests that almost all control subjects, OFC lesion patients and BG lesion patients deleted investing 0 as strong player, yet around 15% of DLPFC patients selected 0 as strong player throughout 80 rounds. Moreover, in the 2nd and 3rd row of the same panel,

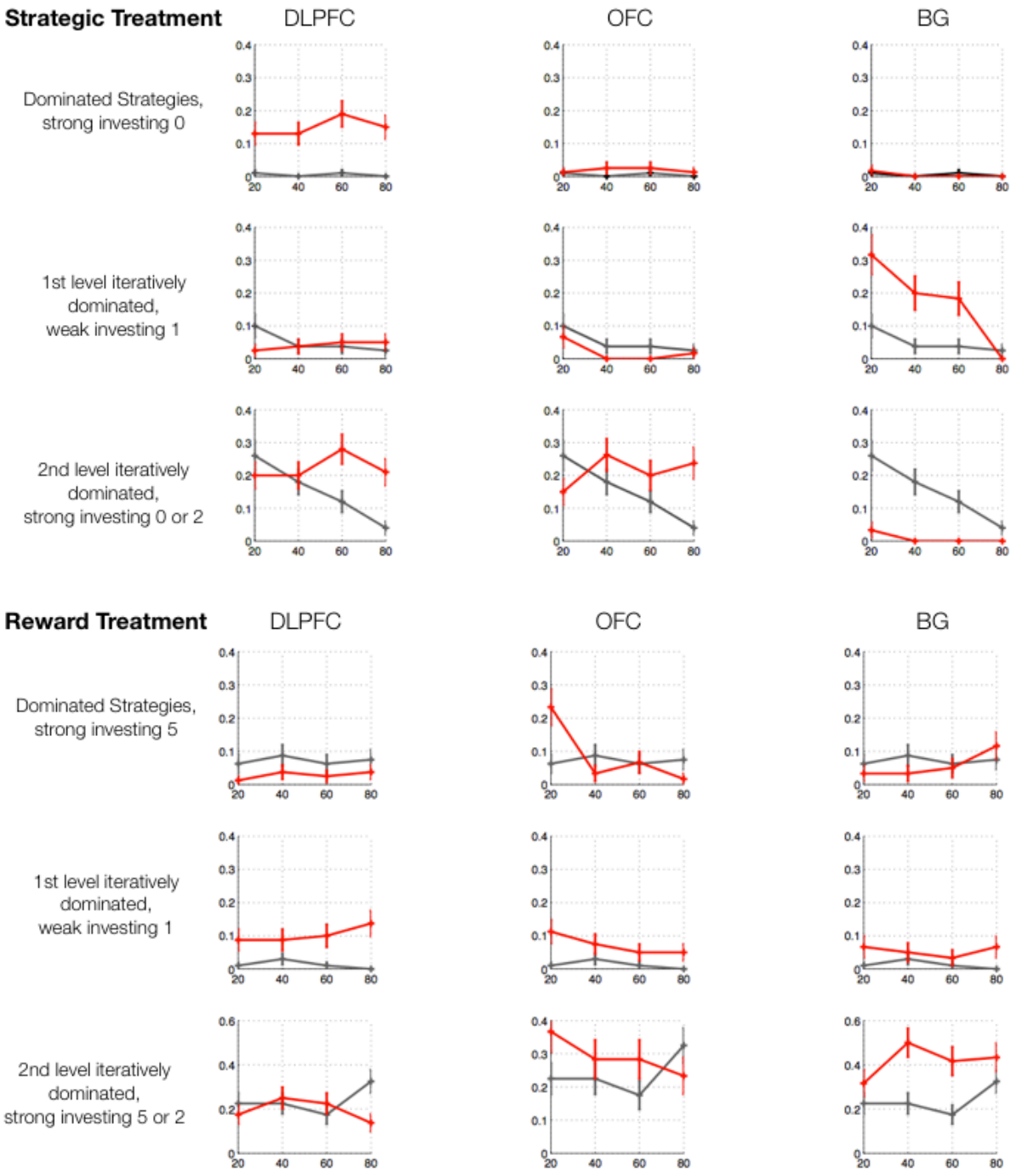

Figure 24: Iteratively elimination of strongly dominated strategies. The empirical cumulative distributions of the portion of lesion subjects playing the corresponding iteratively dominated strategies (red line), compared to the healthy controls(grey line). X-axis indicates the frequency of playing the corresponding strategies. Y-axis represents the portion of experimental cohort. Up: strategic treatment. Down: reward learning treatment.

the healthy controls (grey line) significantly eliminated the usage of 1st and 2nd order of iteratively strongly dominated strategies. Within lesion patients, DLPFC patients showed no sign of reduction of the usage of such strategies, and BG patients surprisingly significantly deleted those strategies over the 80 rounds, whereas OFC patients only significantly eliminated 1st order of iteratively dominated strategy but not the second order one.

Interestingly in the control treatment of reward learning, there exits no significant elimination of iteratively dominated strategies for both healthy controls and lesion subjects, as presented in the bottom panel of Figure 24.

To summarize, these results imply that (1) BG patients appeared to learn to eliminate iteratively dominated strategies in strategic learning. (2) OFC patients seemed to be bale to eliminate the lower order of iteratively dominated strategy but not the higher order one, which is in line with the existing studies suggesting that patients with lesion to OFC region perform well in simple theory of mind test but show deficits inn more advanced one. (3) DLPFC patients seemed to be impaired for such strategic reasoning at all levels.

Model-Based Estimation

This qualitative pattern is supported by our model-based measures (Figure 25). The standard errors of the EWA estimators are estimated with a jackknife approach. More specifically, for a lesion group of sample size n, n runs of EWA estimations were computed. In each run, one subject was excluded from the corresponding lesion group and the EWA model was estimated using the remaining (n-1) subjects. The standard

errors of the EWA estimators are then the standard deviations of jackknife estimates across the N runs. To evaluate how well the model fits the data, we also computed pseudo R^2, which shows how much EWA model outperforms the model of random choice. Pseudo R^2 is defined as the difference between the AIC of EWA and the log likelihood of a random model, scaled by the random-model log likelihood. Hence when the value is equal to 1, it means that the model predicts all choices with probability 1. If its value is equal to 0, it means that the model's explanatory power is the same as a pure random choice model.

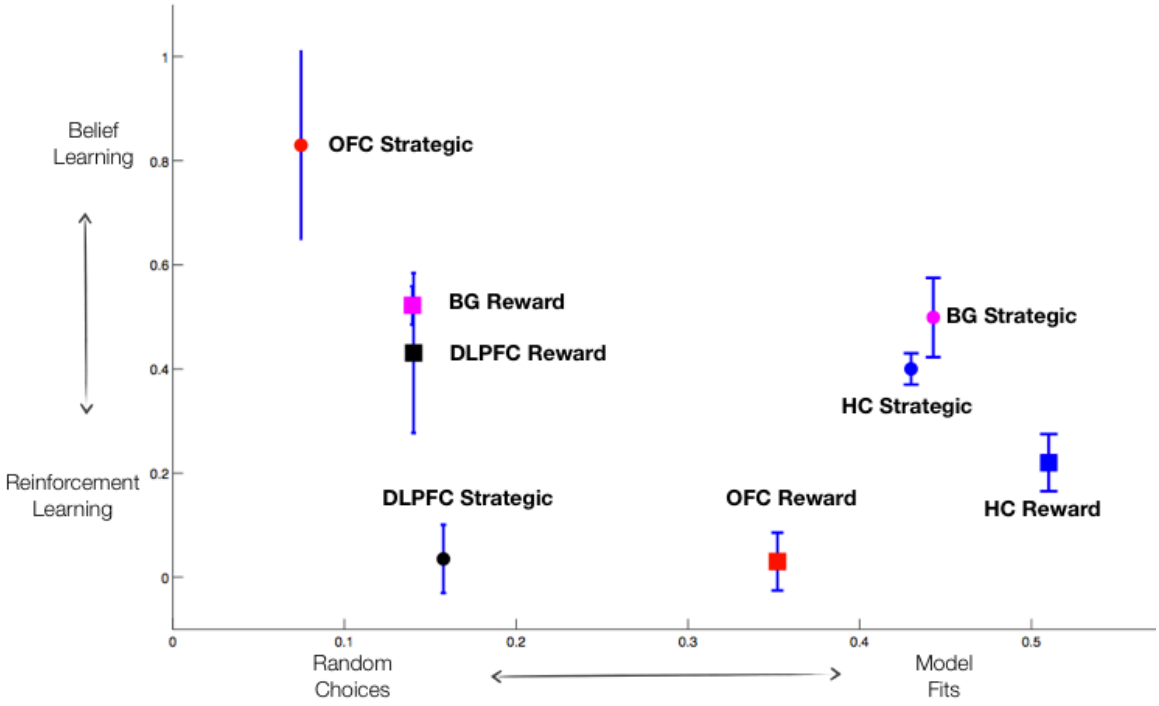

Figure 25: EWA estimation results for different lesion groups. X-axis is pseudo R^2 and thus shows how well our model captures the data relative to a model where players choose randomly. The Y-axis is the estimated value of δ, which measures how much subjects' choices are driven by belief learning relative to reinforcement learning. The x-axis A quantity of 0 thus implies the model fits no better than a random model. Error bars are standard errors for δ, computed with Jackknife method.

Figure 25 visualized the main estimation result. On the x-axis, it is the value of pseudo R^2 for the lesion group under the particular treatment. On the y-axis, it is the estimated value of δ, which captures the weight on the belief based learning relative to reinforcement learning. Figure 9 shows that: (1) choices of DLPFC patients cannot be captured by EWA model well, as indicated by the values of pseudo R^2; (2) Choices of BG reward and OFC strategic are also not well captured by EWA model, i.e. the explanatory power of the EWA is not much higher than a random model; (3) in contrast, data from BG strategic and OFC reward session fit better. (4) Estimated value for δ is higher in the BG strategic cohort than in the OFC reward cohort.

Taken together, these results suggest that (1) the DLPFC patients appear completely impaired in both social and reward learning treatments; (2) it appears to be a double dissociation between the two treatments, strategic and reward leaning, and the brain regions, the BG and OFC. This raises the interesting possibility that the BG patients are able to use the belief learning regions to substitute for the impairments in reward learning deficits.

Model-based estimation and interactively elimination of dominated strategies

To further verify that when our computational model has explanatory power, it was able to quantify the pattern of data apparent in the model-free measurements, we linked the model-based estimation with the iteratively elimination of dominated strategies through simulation. Figure 26 presents the simulated results for reward and strategic learning, which is consistent with our previous findings.

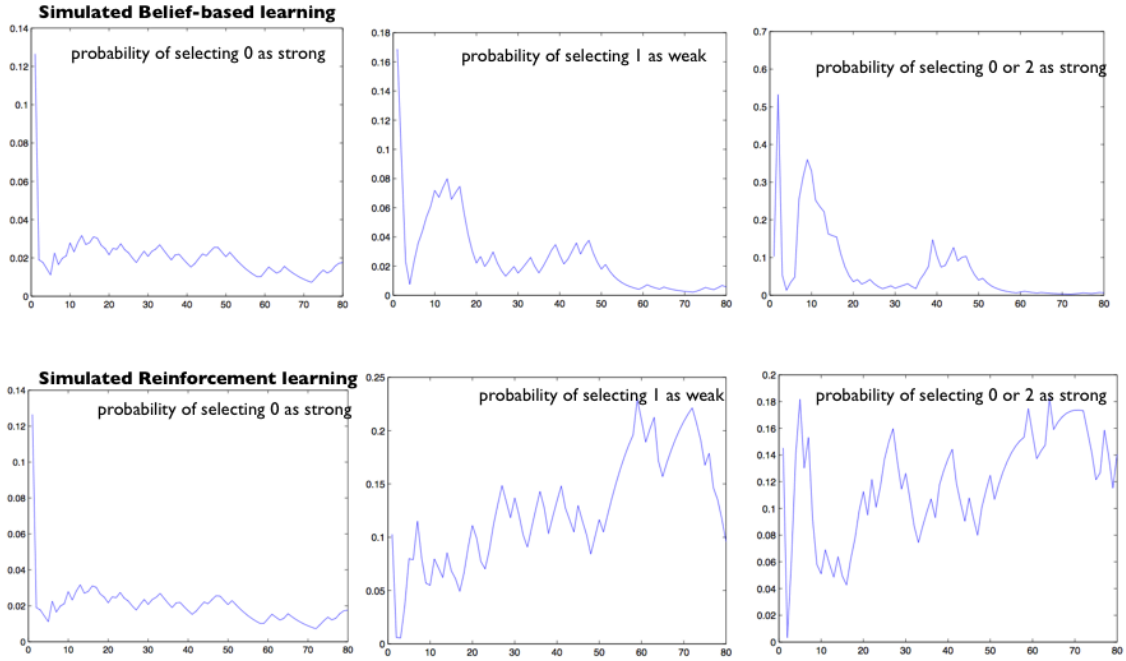

Figure 26: Simulated probability of selecting the interactively dominated strategies under reinforcement and belief-based learning respectively. Up: simulated belie-based learning. Bottom: simulated reinforcement learning.

3.5. Discussions

In this chapter study we aimed at investigating the causal relationships between brain regions to strategic learning. Specifically, we tested whether those regions are necessary for strategic learning, by comparing the behavior of patients with lesion to orbitofrontal cortex (n = 7), basal ganglia (n = 6), and dorsolateral prefrontal cortex (n = 9) across strategic learning and a control treatment of non-strategic learning. Our preliminary result suggests that OFC patients performed well in non-strategic tasks yet showed deficits in strategic treatment, whereas BG patients were impaired in non-strategic learning. Somewhat surprisingly, BG patients performed better in strategic treatment compared to non-strategic treatment, which raises the interesting possibility

that the BG patients are able to use the theory of mind regions to substitute for the impairments in reward learning deficits. Finally we found that DLPFC patients appear completely impaired in both strategic and reward learning treatments. This of course can be due to a number of possibilities. We will accordingly include additional reward and strategic learning tasks used in the previous literature to test the specificity of the effects across tasks in this cohort, e.g. Iowa Gambling Task, Wisconsin Card Sort, and the multi-round Trust Game.

Overall, the BG and OFC results provide preliminary support for our hypothesis that these regions are respectively necessary for reinforcement and belief-based learning that subserved both reward and strategic learning. On the other hand, the poor fit of the model in some of the treatments calls for a revision of our model in order to provide a model-based measure of the impaired behavior and functions and not merely of the preserved behavior and functioning. Finally, we will in our future work to study the learning behavior at individual level and associate the symptom severity with both individualized model free and model-based measurements.

3.6. Patient MRIs and individual transition matrices

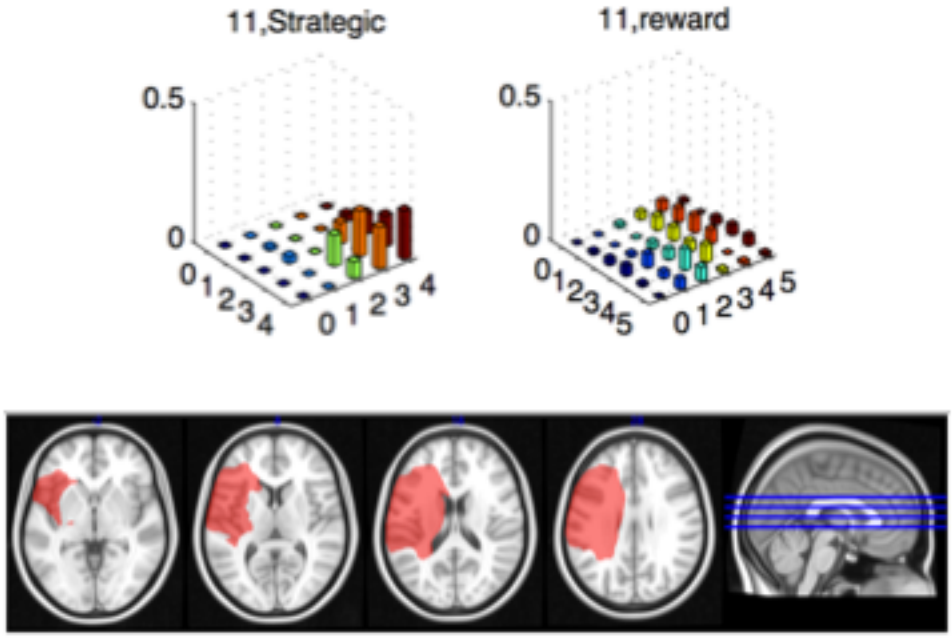

Figure 27: UP: Transition matrices for both strategic and reward treatment for subject 11. Down: Lesion area for subject 11.

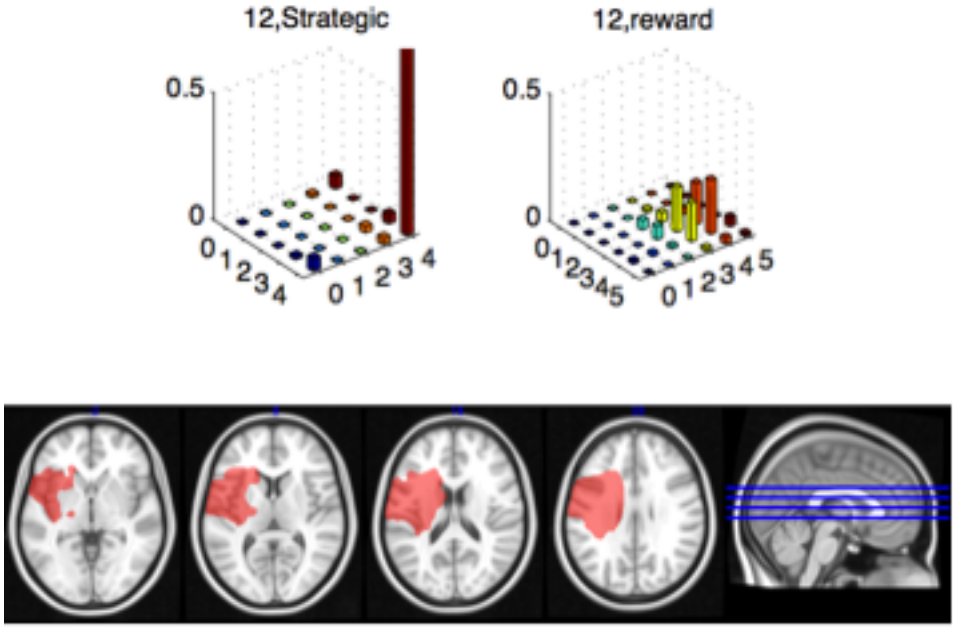

Figure 28: UP: Transition matrices for both strategic and reward treatment for subject 12. Down: Lesion area for subject 12.

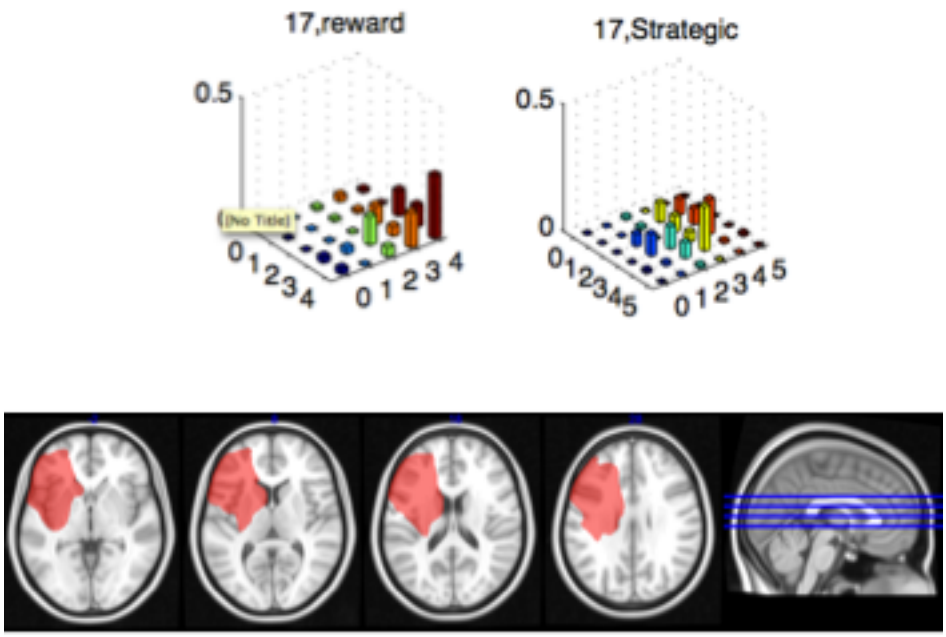

Figure 29: UP: Transition matrices for both strategic and reward treatment for subject 17. Down: Lesion area for subject 17.

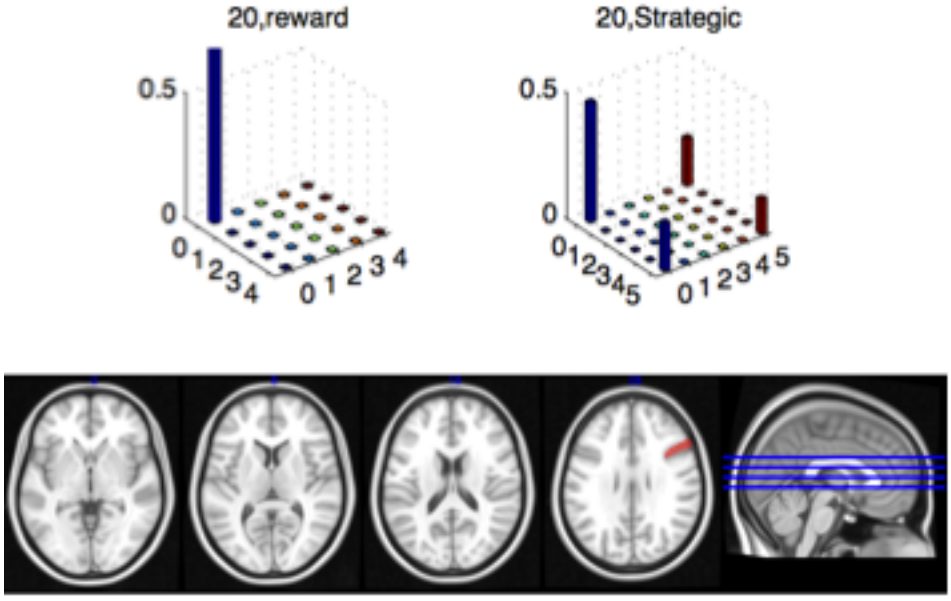

Figure 30: UP: Transition matrices for both strategic and reward treatment for subject 20. Down: Lesion area for subject 20.

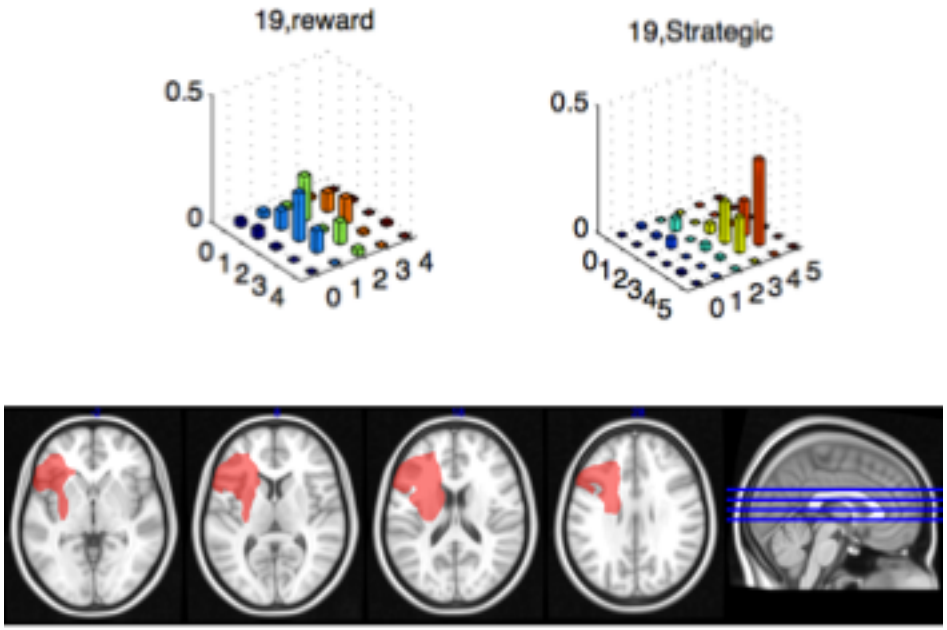

Figure 31: UP: Transition matrices for both strategic and reward treatment for subject 19. Down: Lesion area for subject 19.

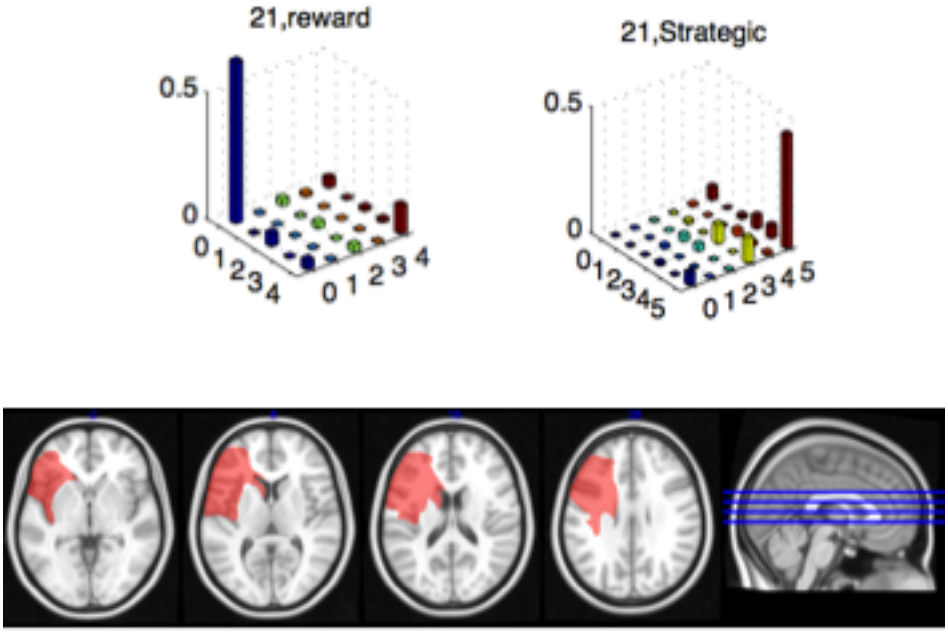

Figure 32: UP: Transition matrices for both strategic and reward treatment for subject 21. Down: Lesion area for subject 21.

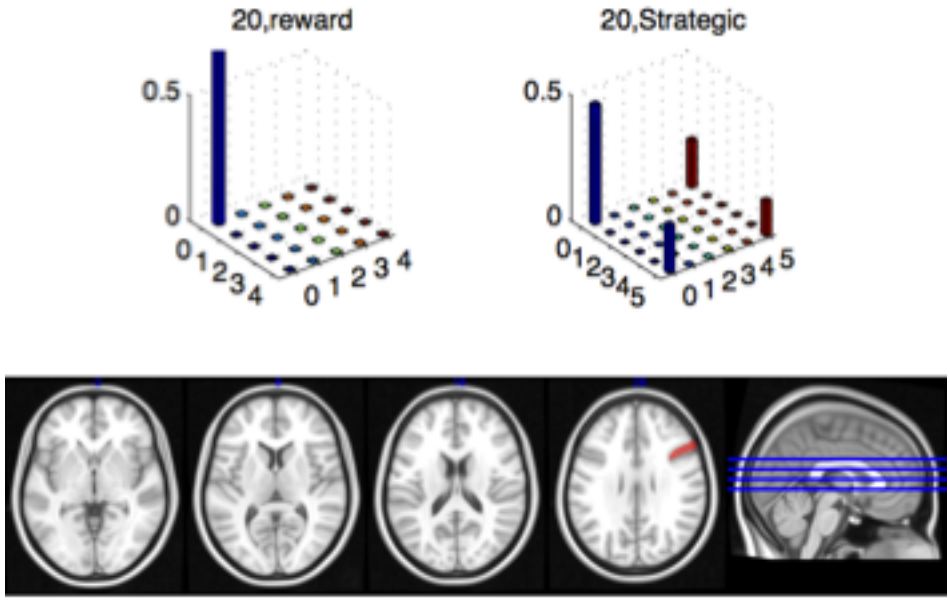

Figure 33: UP: Transition matrices for both strategic and reward treatment for subject 20. Down: Lesion area for subject 20.

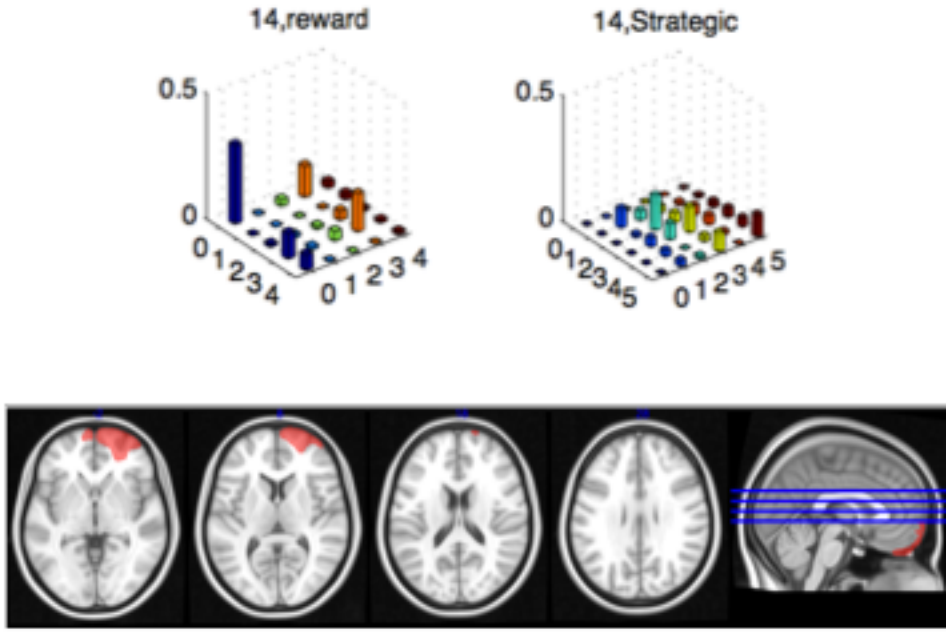

Figure 34: UP: Transition matrices for both strategic and reward treatment for subject 14. Down: Lesion area for subject 14.

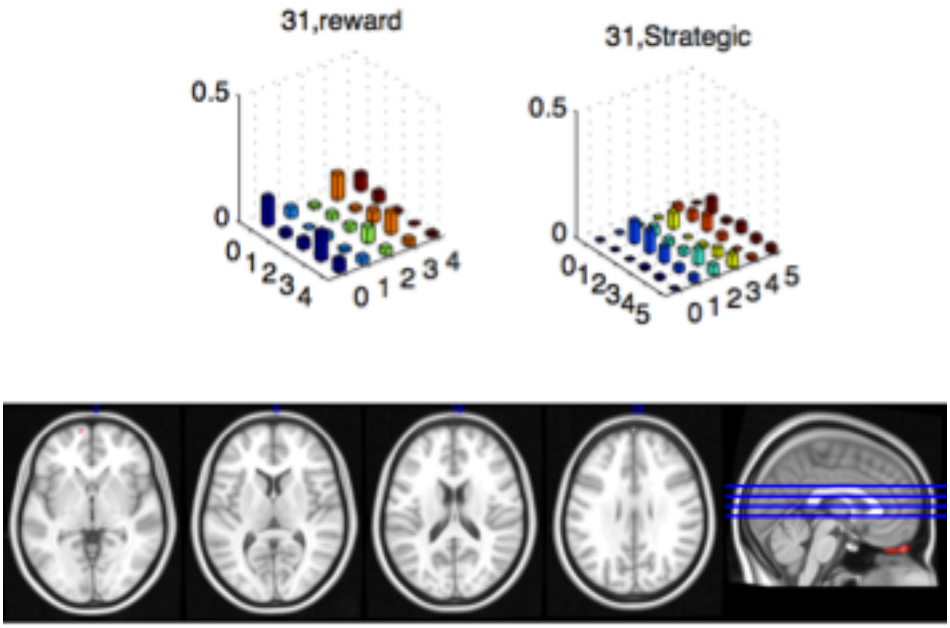

Figure 35: UP: Transition matrices for both strategic and reward treatment for subject 31. Down: Lesion area for subject 31.

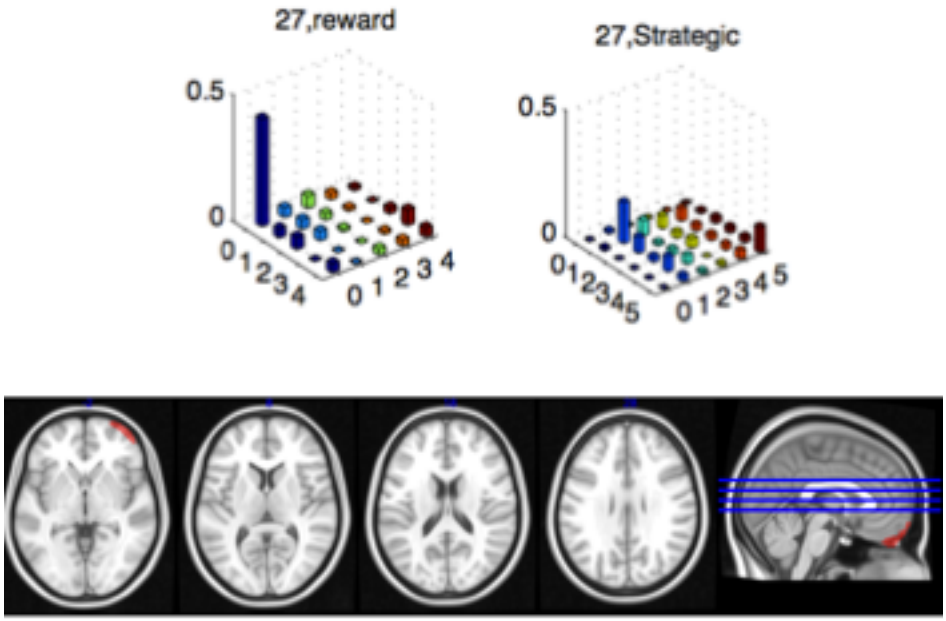

Figure 36: UP: Transition matrices for both strategic and reward treatment for subject 27. Down: Lesion area for subject 27.

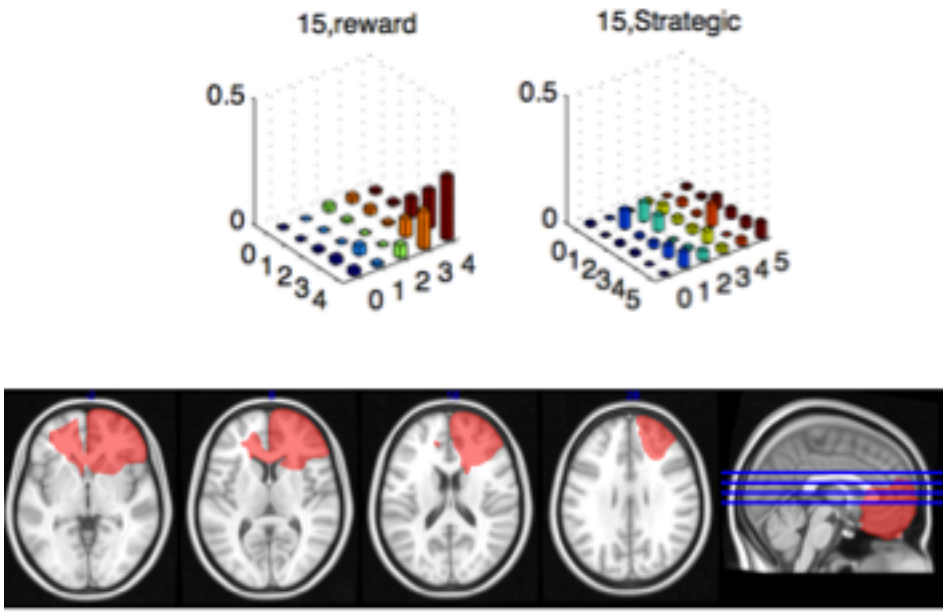

Figure 37: UP: Transition matrices for both strategic and reward treatment for subject 15. Down: Lesion area for subject 15.

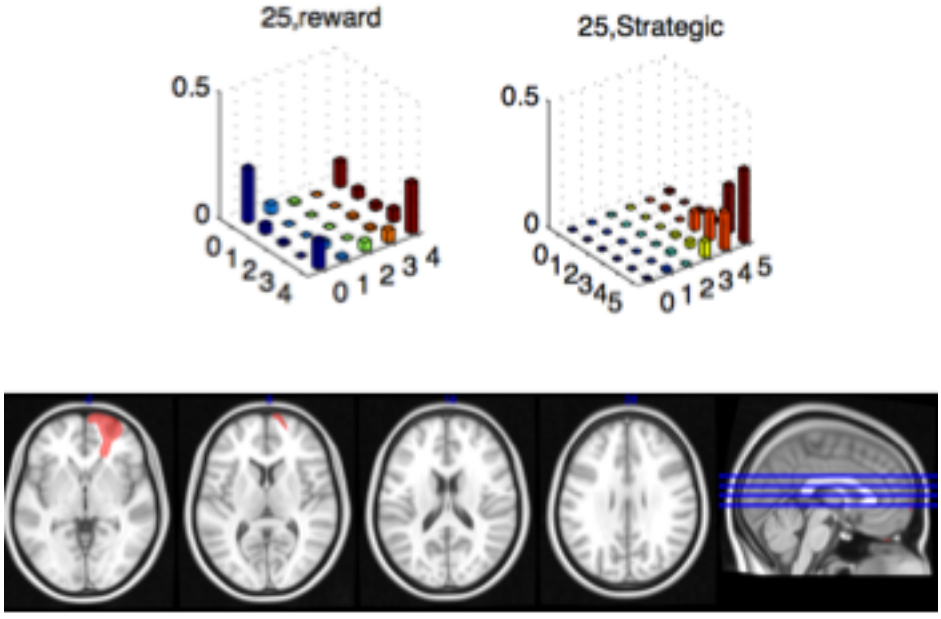

Figure 38: UP: Transition matrices for both strategic and reward treatment for subject 25. Down: Lesion area for subject 25.

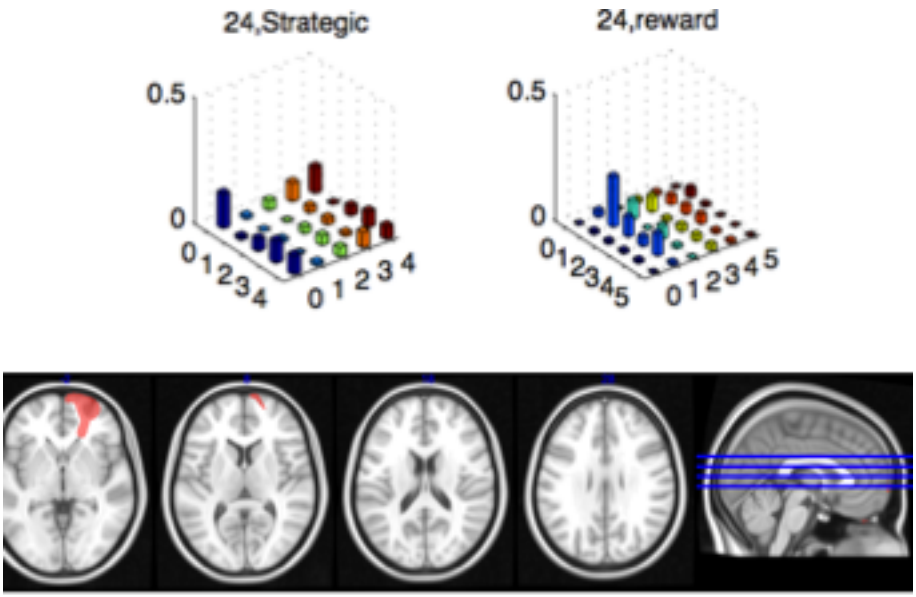

Figure 39: UP: Transition matrices for both strategic and reward treatment for subject 24. Down: Lesion area for subject 24.

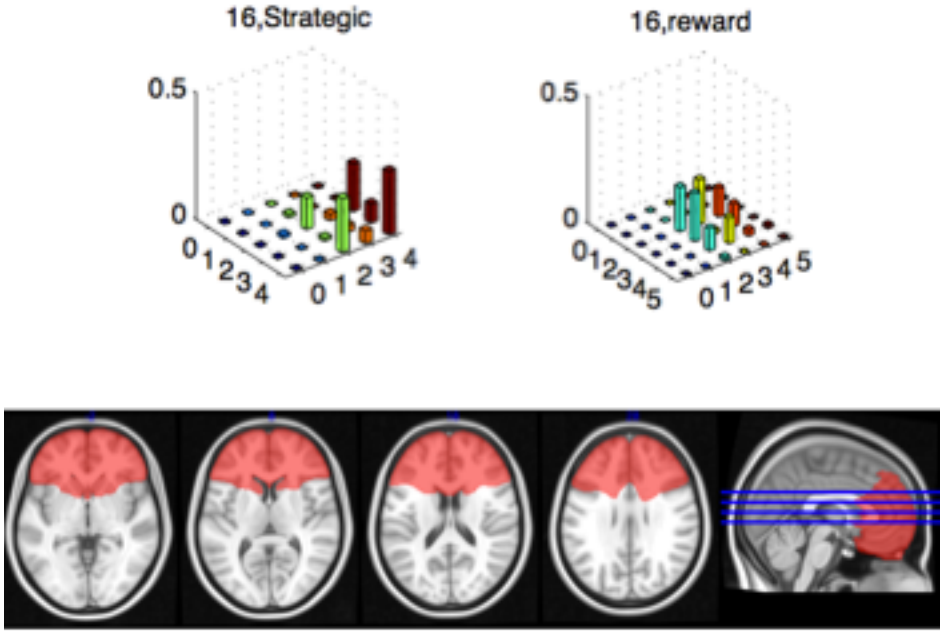

Figure 40: UP: Transition matrices for both strategic and reward treatment for subject 16. Down: Lesion area for subject 16.

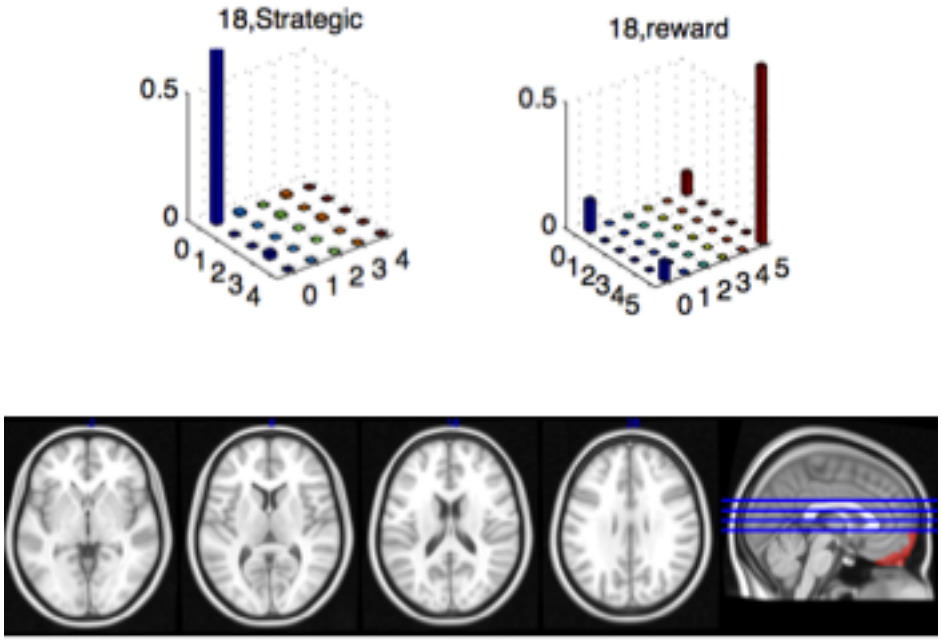

Figure 41: UP: Transition matrices for both strategic and reward treatment for subject 18. Down: Lesion area for subject 18.

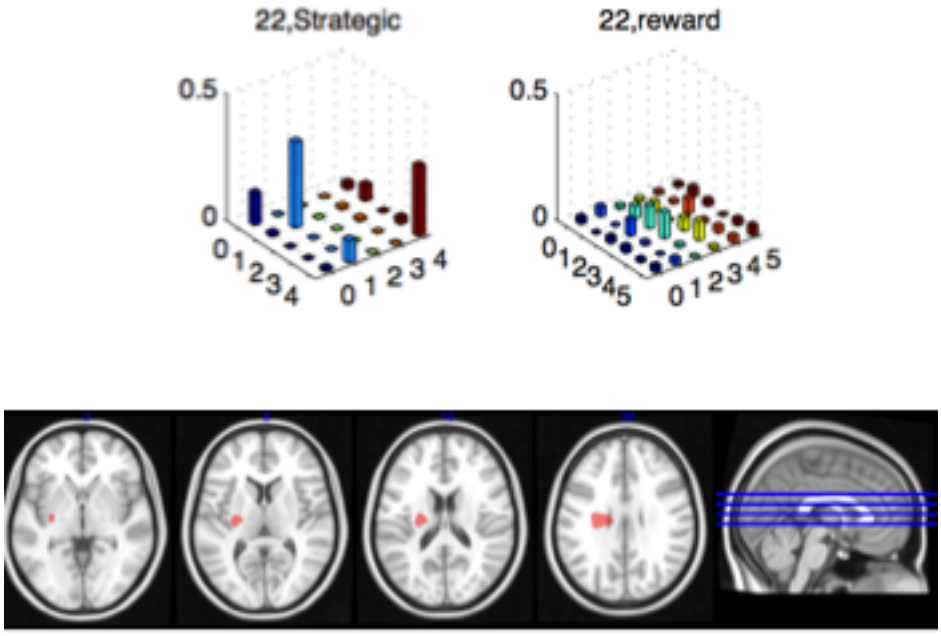

Figure 42: UP: Transition matrices for both strategic and reward treatment for subject 22. Down: Lesion area for subject 22.

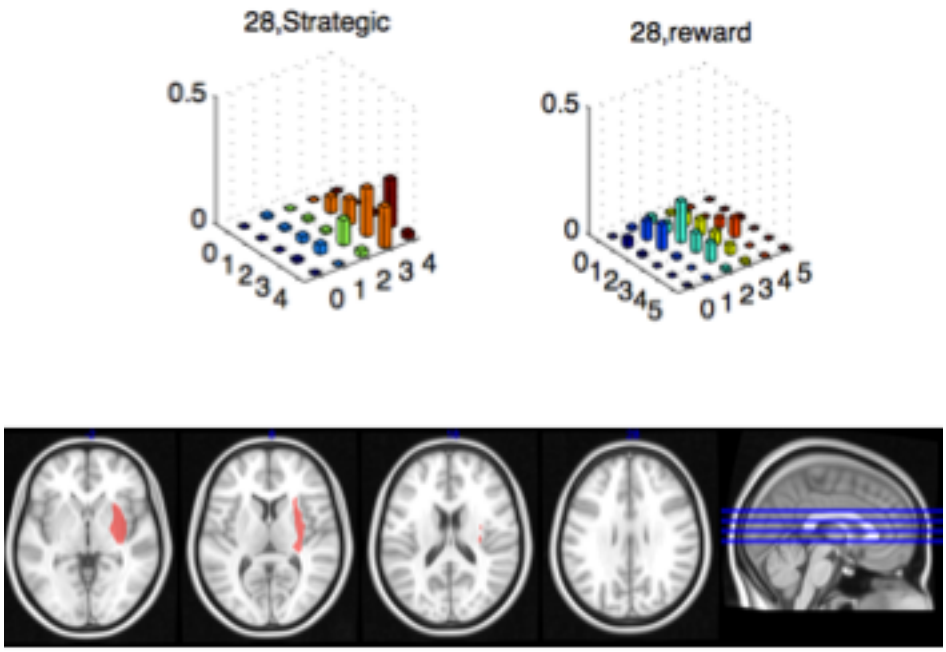

Figure 43: UP: Transition matrices for both strategic and reward treatment for subject 28. Down: Lesion area for subject 28.

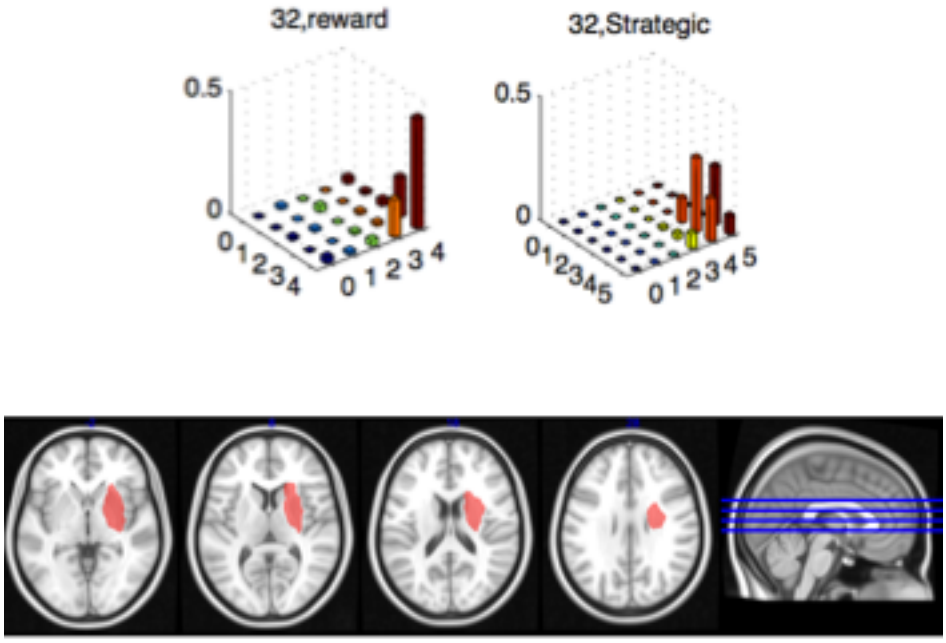

Figure 44: UP: Transition matrices for both strategic and reward treatment for subject 32. Down: Lesion area for subject 32.

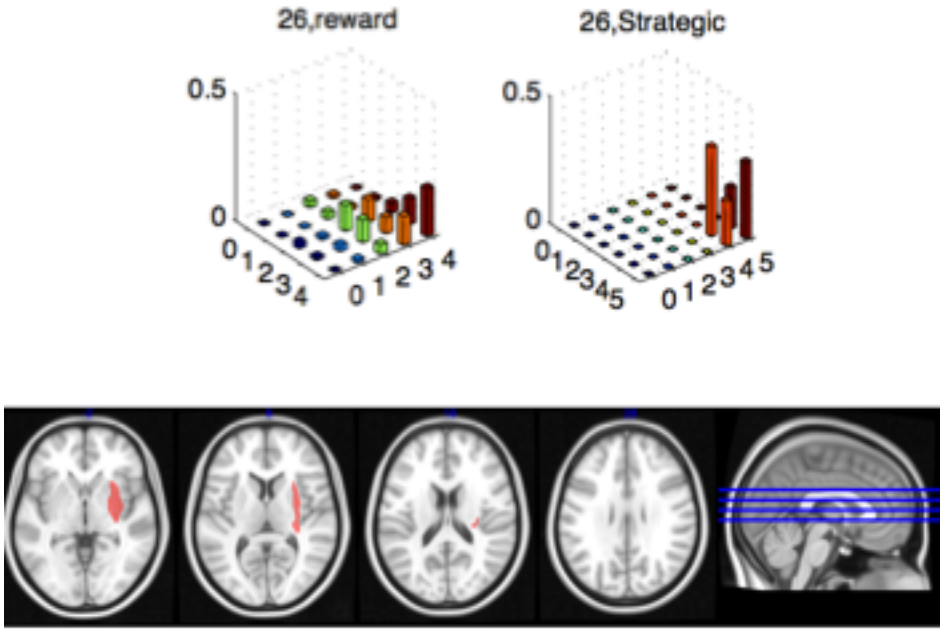

Figure 45: UP: Transition matrices for both strategic and reward treatment for subject 26. Down: Lesion area for subject 26.

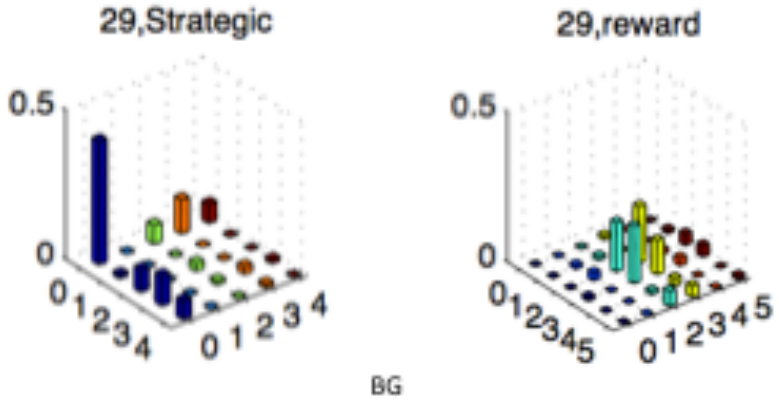

Figure 46: Transition matrices for both strategic and reward treatment for subject 29. Subject 29 is known to have lesion in Basal Ganglia. MRI is unavailable.

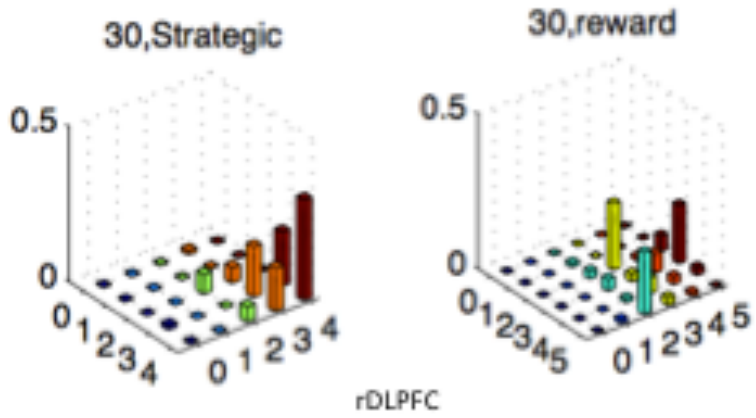

Figure 47: Transition matrices for both strategic and reward treatment for subject 30. Subject 30 is known to have lesion in right DLPFC. MRI is unavailable.

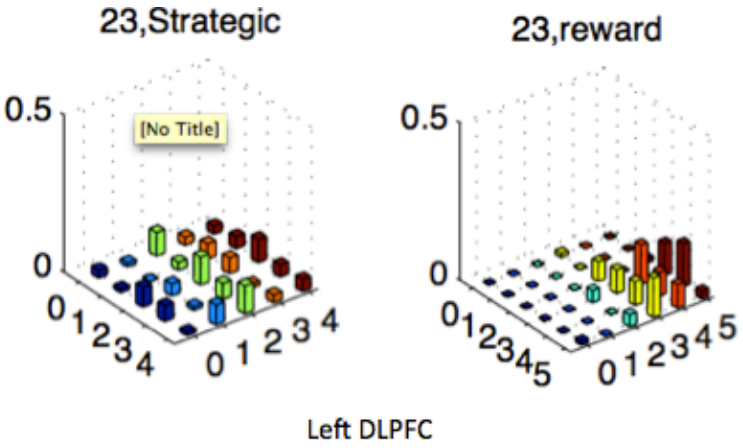

Figure 48: Transition matrices for both strategic and reward treatment for subject 23. Subject 23 is known to have lesion in left DLPFC. MRI is unavailable.

3.7. Reaction time for each lesion groups across treatments.

Reaction time is defined as the time from the game presentation onset to the end of the decision-making, which is widely used in lesion studies for both human and non-human primates (Frank, Seeberger et al. 2004; Frank, Samanta et al. 2007; Buckley, Mansouri et al. 2009). Although we think it is only indirect information associated with strategic learning, we still included the median reaction time for each lesion group under each treatment in Figure 49. Note that according to Figure 49, the BG patients have the most distinct pattern of reaction time, i.e. they are slower in reward learning than in the strategic learning, which may be interpreted as a supporting evidence for our model-based estimation result that BG patients are less impaired in strategic learning.

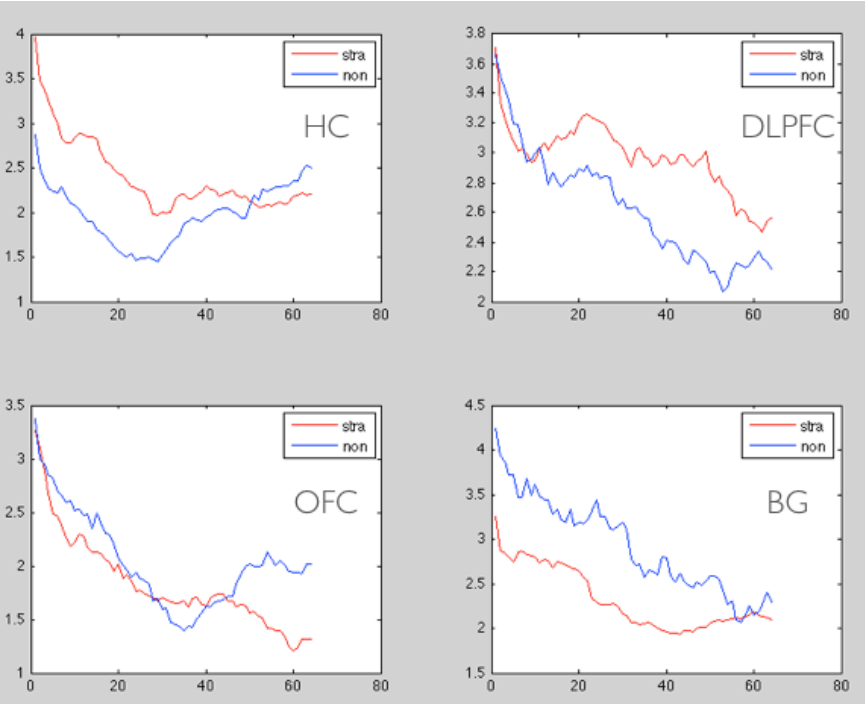

Figure 49: Reaction time for each lesion group under each treatment. The median action times for each lesion group and are smoothed over a time window of 15 periods.

CHAPTER 4

AGING AND STRATEGIC LEARNING

4.1. Introduction

It has been widely recognized that the portion of the elderly adults in the global population will keep growing, due to the increase in average life expectancies as well as the decline in birth rates (United Nation and Economic Division 2001). As a result, the elderly's economic decisions have gained higher relative social impacts. To take just one measure, the median net worth of a 65-year-old American in 2007 is more than double that of a 40-year-old (Bucks, Kennickell et al. 2009). Despite the growing demand for understanding how economic decision-making change over the lifespan, only a handful researches have focused on characterizing the relationship between economic decision and aging at behavioral level (Kovalchik, Camerer et al. 2005), and identifying how functional, anatomical, and neurochemical changes in the aging brain will affect the competence of economic decision making (Li, Biele et al. 2007; Samanez-Larkin, Gibbs et al. 2007; Samanez-Larkin, Kuhnen et al. 2010).

Some popular stereotypes suggest that the ability for making strategic decisions declines with aging, yet other commonly held beliefs emphasize the wisdom of age especially for tasks involved pragmatic reasoning (Grossmann, Na et al. 2010); (Staudinger, Cornelius et al. 1989; Nyberg, Sandblom et al. 2003). Real world empirical evidences have suggested that the elderly population is particularly vulnerable in a number of ways. For example, the elderly are disproportionate targets of fraud across the world (Templeton and Kirkman 2007). They are thought to constitute approximate 30%

of all fraud victims in the United States, but even this figure is likely to be too low due to likelihood of widespread underreporting (Bucks et al. 2009). It has been hypothesized that such vulnerability can be in part be attributed to a decline in the ability for strategic thinking in order to avoid being exploited by others. Similarly, research has documented that the elderly investors continued to invest in stock markets even after losses large enough to postpone their retirement, suggesting a relatively flat learning curve when adapting to an uncertain environment (Rappaport and Dragut 2005).

Yet it is challenging to identify the precise deficits underlying such vulnerability at the behavioral and neural levels. Unlike memory and motor impairments, which are readily recognized as symptoms of more serious underlying neurological conditions, decision-making deficits often do not elicit comparable concern in the elderly (Denburg, Tranel et al. 2005). There are also few neuropsychological tools or biomarkers available to measure and quantify decision-making deficits, particularly those that contain a social component.

In this paper I studied a specific type of economic decision for the elderly sample, the strategic learning in a repeated normal form game. More specifically, I examined the elderly's decision making (1) in an environment where outcomes do not depend on his or her own choice alone but upon the choices of others; and (2) from a dynamic point of view when choices unfold over time. That is I studied how the elderly learns from the past rewards and punishments as well as the possible strategies used by other intellectual agents in order to avoid being exploited by others. Such decision-making is ubiquitous in social life. Examples include playing poker, investing in stock market, or trade negotiation. Understanding how the ability for strategic learning changes over the adult

lifespan will be have important social and economic implications, such as providing individual cognitive improvement as well as guide for social policy making, ranging from allocating personal wealth to selecting medical care.

4.2. Literature Review

Recent behavioral cognitive researches suggested variable degeneration across different cognitive abilities over adult lifespan (Figure 50; Hedden and Gabrieli 2004; Reuter-Lorenz and Lustig 2005; Dennis and Cabeza 2008; Grady 2008). For example, working memory and perceptual speed were found robustly decline with aging, while semantic knowledge and emotional processing remain stable (Schaie 1996; Walhovd, Fjell et al. 2005; Williams, Brown et al. 2006). Similarly, different patterns of economic behaviors have also been mapped across different age groups, such as changes in social preferences (Kovalchik, Camerer et al. 2005), risk attitude (Kovalchik, Camerer et al. 2005; Samanez-Larkin, Kuhnen et al. 2010), and discounting rate (Green, Fry et al. 1994). These findings suggest distinct age-related affect on neural substrates underlying different functions (Hedden and Gabrieli 2004). Here I review the literature related to age-associated change in strategic learning from two aspects: strategic reasoning (theory of mind) and learning.

Aging and Theory of Mind

The ability of guessing what others will do without any prior contractual agreements and use this knowledge in social interaction is referred to “theory of mind”

(Happé, Winner et al. 1998; Amodio and Frith 2006) or strategic reasoning (Camerer, Ho et al. 2003), which has been extensively studied from both psychology and economics

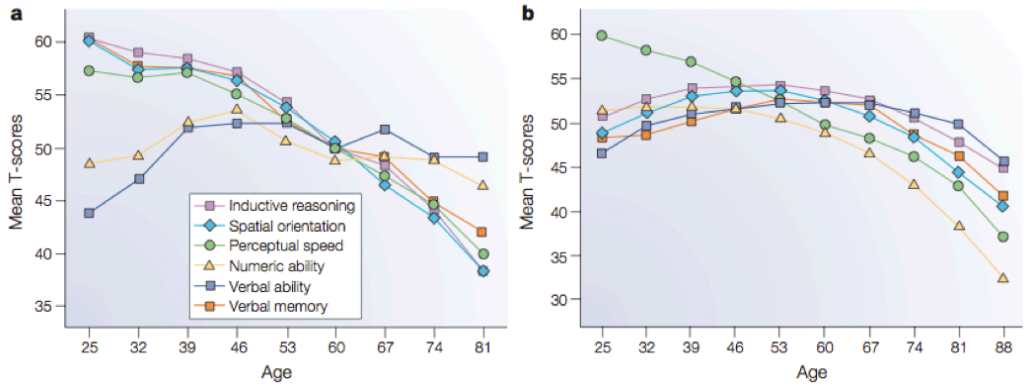

Figure 50: (A) cross-sectional and (B) longitudinal studies on the changes in a variety of cognitive functions across adult lifespan. Adapted from Schaie (1996)

perspectives. In psychology, verbal and graphical stories and are often used to examine subjects' ability of making inferences about the thoughts and feeling of the characters (Amodio and Frith 2006). For example, some (Charlton, Barrick et al. 2009; McKinnon and Moscovitch 2007; Slessor, Phillips et al. 2007) suggested a decline in the ability of theory of mind with normal aging. Yet other studies (Happé, Winner et al. 1998); (Grossmann, Na et al. 2010) found that older adults actually performed better than younger adults, in the face of possible decline in many forms of cognitive processing.

In economics, game theory paradigm is one of the most popular tools for assessing the ability of strategic reasoning. Traditionally, notions of equilibria have been developed based on the assumption that players have perfect theory of mind and can guess accurately what others would do. While the concept of equilibrium are useful as an idealized model, a large number of experimental studies have shown that actual subjects

do not always make equilibrium choices and exhibit significant amount of individual differences in the level of theory of mind. Recent studies have further suggested that

Stem and Leaf Plots for the Beauty Contest Game

Younger Subjects		Older Subjects
0	47	
1	5679	4788
2	000023478889	12255677779
3	022333345555556677788	2355557
4	222235	457
5	0002	0028
6	58	25
7		5
8		6

Figure 51: Healthy elderly subjects (average age 82) manifested no significant difference from young subjects (average age 20) in the beauty-contest game. In the plot, the left column is the tens digits, whereas the middle and left columns are the digits selected by each cohort. For instance, values within 30 to 39 have been selected by the young cohort most frequently. Similarly the elderly subjects picked values from 20 to 29 most frequently. Overall, both young and elderly cohort clustered at level 2 and 3 of strategic reasoning, yet the elderly is slightly more likely to select numbers larger than 50 ($p<0.07$). Adapted from Kovalchik, Camerer et al. (2005)

there may exist reliable behavioral measurement on abilities reasoning (Stahl 1996; Camerer, Ho et al. 2004; Heinemann, Nagel et al. 2009) and its biology basis (Coricelli and Nagel 2009; Haruno and Kawato 2009; Bhatt, Lohrenz et al. 2010) based on young population.

Kovalchik et al. (Kovalchik, Camerer et al. 2005) compared the ability of strategic reasoning between the young and healthy elderly subjects in a "p-beauty contest" paradigm, which has been widely used in many behavioral studies over a variety of representative populations, including business executives and collage students (Nagel

1995; Duffy and Nagel 1997). In Kovalchik's task, subjects were told to play with nine other individuals from their research cohort by selecting any integer from 0 to 100. The winner is the individual who is closest to 2/3 of the average of all the numbers selected in his/her cohort. Under such a setting, naive players would select answers randomly, resulting in a uniformly distributed number with an average of 50. Conditional on the belief that others are random players, one will select 2/3 of 50. Similarly, if one believes that all others reason their way to select 2/3 of 50, his/her best response will be (2/3)*(2/3) of 50. Such iterative reasoning process could go on and on depending on subjects' belief about what his/her group members will do. Kovalchik found (Figure 51) no significant difference between the healthy elderly and young subjects, with both clustering at the second and third level of strategic reasoning.

Aging and Learning

Social interactions are rarely one-short games. Much of the evidence from experimental economics has suggested significant dynamic changes in strategic reasoning during repeated games (e.g. (Huck, Normann et al. 1999). However, these behavioral studies are mostly based on samples from young college students.

Recent studies from cognitive neuroscience have compared reward learning between younger and older adults at both behavioral and neural level, suggesting the elderly adults has a reduced ability in accurately predicting the expected reward (Marschner, Mell et al. 2005; Schott, Niehaus et al. 2007), or learning from positive outcome (Mell, Heekeren et al. 2005).

Another popular task for assessing subjects' ability to identify favorable choices in the long run is the so-call "Iowa Gambling Task" (IGT). In IGT (Bechara, Tranel et al. 1997; Maia and McClelland 2004) subjects choose between 4 decks of cards repeatedly. Deck A and B are composed of high returns and high losses, and Deck C and D with low returns and low losses. Overall Deck C and D have negative mean with higher variance whereas Deck A and B have positive mean with lower variance. Moreover, decks are manipulated in such a way that subjects usually start with favoring the "bad decks", Deck A and B, which yield high payoffs in the first a few rounds,

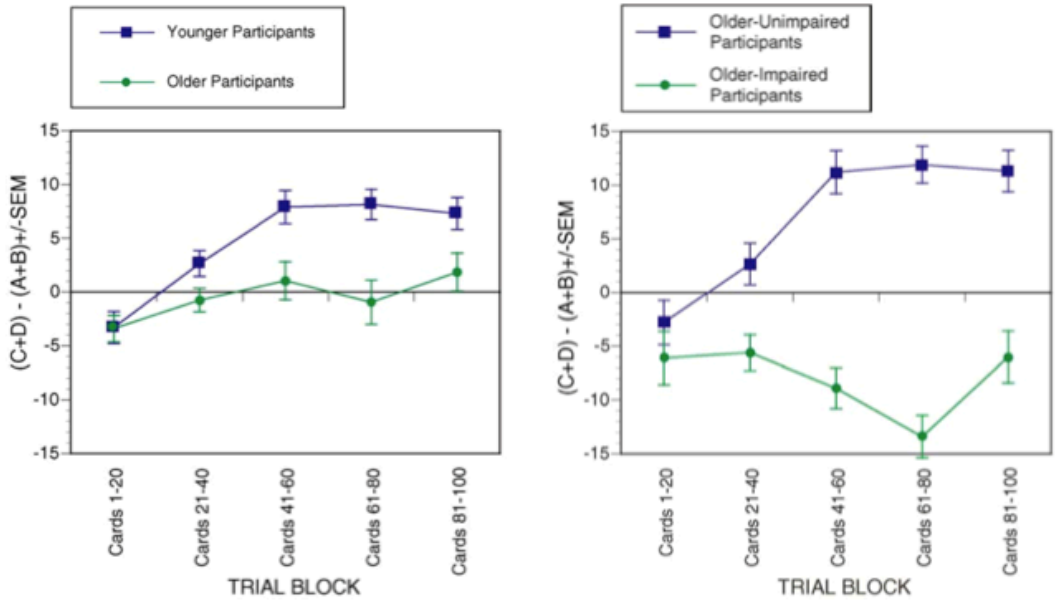

Figure 52: A subset of the neurologically and psychiatrically healthy elder group of subjects demonstrated flatter learning curve during the Iowa Gambling Task, in spite of otherwise intact cognitive functioning." Adapted from Denburg (2005, 2007)

and then switch to Deck C and D after they learn about the associated average payoff and variances from repeated sampling.

A series of studies from the University of Iowa reported that at least a subset of the elderly subjects were locked in with the "bad" decks, indicating an impaired ability to identify favorable options in the long run (Denburg, Tranel et al. 2005; Denburg, Cole et al. 2007). Yet Kovalchik, Camerer et al. (2005) conducted a slightly different version of IGT (with only two decks) and found no significant difference between the young and elderly populations.

From a formal perspective, IGT is somewhat similar to the well-studied multiple-armed bandit problem in decision theory and dynamic control, but with poorly defined information set. Subjects have no information about the range or stationary of the underlying probabilities for each deck. Moreover, the draws are manipulated and are done without replacement. Consequently, a subject who failed to move to the "good" deck could totally because of ambiguity-aversion rather than impaired ability of learning.

4.3. Hypotheses

My hypotheses about age-associated change in strategic learning are based on (1) my previous studies on the neural computation underlying strategic learning, and (2) well-established neural evidences on aging brain. My previous research suggests that strategic learning straddles both reward and belief based learning at both behavioral and neural level, with striatal area involved in both reward reinforcement and belief based learning, and medial prefrontal region associated only with belief-based learning.

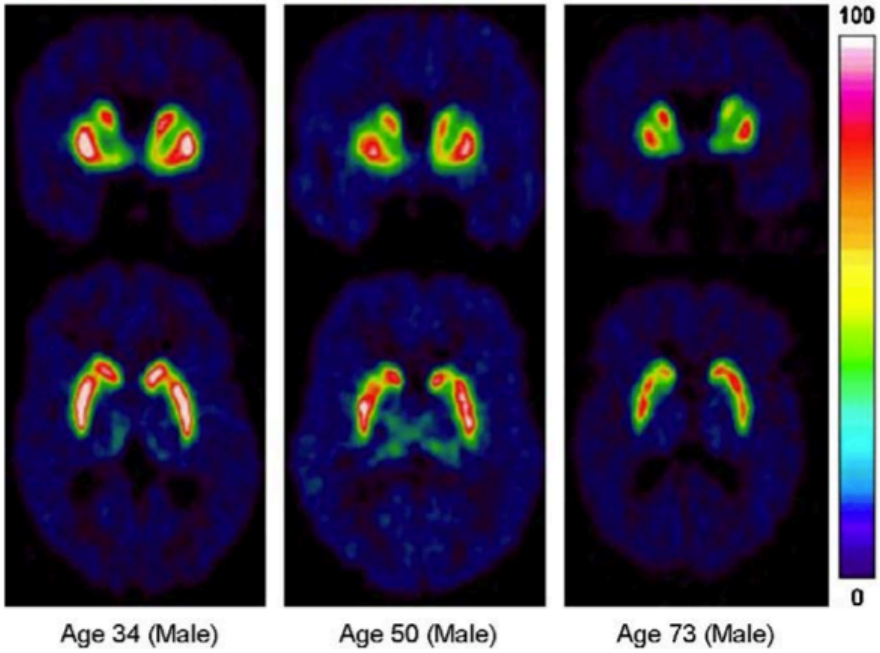

Figure 53: Age-associated loss of dopamine transporter density in the striatum (caudate and putamen) from PET studies. Adapted from Samanez-Larkin, Gibbs et al. (2007).

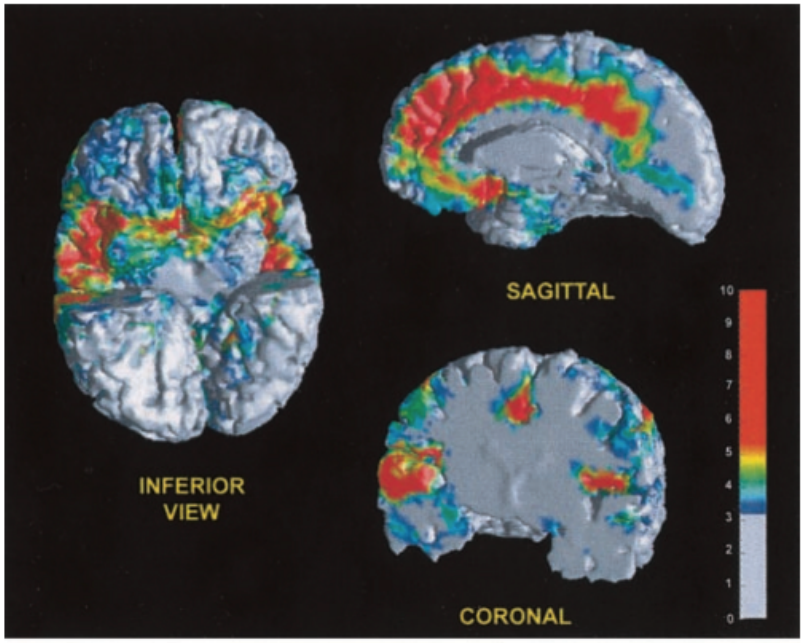

Figure 54: Significant tissue loss in gray matter volume in the insula), orbital frontal cortex, and cingulate cortex, from longitudinal studies through MRI. Adapted from Resnick, Pham et al. (2003)

On the other hand, there exist much evidences suggesting degeneration in these brain regions with normal aging. Firstly, dopaminergic system degenerates with normal aging, as shown in Figure 53. There are several lines of evidence supporting the idea that aging can have deleterious effects on decision-making abilities (Marschner, Mell et al. 2005). Among the most persuasive evidence is the crucial role the nigrostriatal dopamininergic (DA) system plays in decision-making and its degeneration during aging. There exists much evidence that an intact dopaminergic system is necessary for reward processing (Knutson, Momenan et al. 2001; Hsu, Bhatt et al. 2005), reward learning (Fiorillo, Tobler et al. 2003), and decisions under both certainty and uncertainty (Schultz, Dayan et al. 1997; Bechara 2000; and Schultz 2000). During aging, there is a decrease in the number of and the number of synapses in these neurons (Erixonlindroth, Farde et al. 2005), possibly underlying abnormalities observed in healthy older adults (Samanez-Larkin, Gibbs et al. 2007).

Secondly, in addition to degeneration of the dopaminergic system, longitudinal studies found frontal lobes suffered the most drastic loss of volume as assessed through MRI (Figure 54, Resnick, Pham et al. 2003). Executive function, as it is commonly conceived, consists of two aspects: an evaluative aspect, related to forming, maintaining, and updating appropriate models of the environment (which may be carried out through various types of memory processes) and an action-oriented aspect, which is instead involved with the coordination of other cognitive functions, including perception, attention, and action. This coordination presumably takes place over time, and is reflected in future behavior, so that when performed appropriately it can lead to successful adaptation to changing task demands.

Combining the above neural evidences together with my previous studies, I hypothesize that at behavioral level: (1) Old adults will perform poorer in Patent Race Game compared to young adults in overall performance; (2) Old adults will learn poorer to reason strategically in Patent Race Game compared to young adults; (3) this difference can be captured by key parameters in my computational model across cohorts.

4.4. Method

Subjects

We compare results from 30 young subjects from University of Illinois at Urbana and Champaign, and 29 elderly subjects recruited from: 1) local flyers and bulletins in the Berkeley community, 2) online forums such as Craigslist, and 3) Berkeley Retirement Center (BRC). Because the BRC serves retired faculty, staff and their families from UC Berkeley and Lawrence Berkeley National Labs, we anticipate the old adult sample will be more highly educated and healthy than the general population.

All elderly subjects are confirmed to be healthy and with no significant neurological issues.

Procedure

Eligible participants were given the Patent Race task, and received a cognitive test battery, including Shipley test for IQ, Bower test for working memory, Wisconsin Card sort test, and RAPN for fluid IQ. At this and all subsequent stages of the experiment subjects were fully informed of the purposes of the research and were free to withdraw without penalty.

A similar protocol was used in the . Upon arrival at the laboratory, subjects were given instructions and quiz to ensure the understanding of the experiment. Participants played two stages of 80 rounds each of strong and weak roles (counterbalanced). Opponents' choices were be drawn from a pool of 16 young adults who participated in an earlier session at University of Illinois at Urbana and Champaign. We did no run all subjects simultaneously because we want all subjects to play against a common distribution of opponents. Previous sessions comparing "live" sessions and "non-live" sessions show that young adults do not differ significantly across treatments**.**

4.5. Model-Free Measurements

Summary Statistics

I started with simple visualization and comparison of choice behavior across age groups. First I compared the empirical frequencies of choices for strong and weak roles across trials and subjects for each age group (Table 7). In order to provide a benchmark for such comparison, the probabilities of the choices predicted by the unique mixed strategy Nash equilibrium are also included. As shown in Table 7, young subjects on average were reasonably close to the Nash equilibrium prediction with the exception of overinvesting 4 and underinvesting 3 as strong players, whereas the distribution of choices made by elderly strong subjects were further away from Nash equilibrium prediction, with more evenly distributed choice over investing 2, 3, 4 and 5. Yet as weak players, both elderly and young subjects overinvested 0 and underinvested 4. However, elderly subjects also overinvested 1, which is the iteratively dominated strategy for the weak role.

Age Group	Young	Elderly
Age	23.3	64.1
	(4.6)	(5.4)
Gender	54%	60%
(% of female)		
Years of Education	14.4	15.0
	(1.1)	(0.9)

Table 6: Demographic information and performance measure across age groups. Means (SEM).

			Empirical Distribution	
Role	Investment	Equilibrium Prediction	Young	Elderly
Strong	0	0%	0.83%	0.67%
	1	20%	17.50%	11.92%
	2	0%	9.54%	15.67%
	3	20%	11.13%	28.65%
	4	0%	16.21%	22.02%
	5	40%	44.79%	21.06%
Weak	0	20%	49.29%	39.32%
	1	0%	3.46%	12.61%
	2	20%	6.46%	7.39%
	3	0%	13.46%	11.14%
	4	40%	27.33%	29.55%

Table 7: Comparison of Nash equilibrium prediction with the Empirical frequencies from young and elderly cohorts.

Second, in order to examine the "stickiness" of choices between successive trials, I computed the average frequencies of repeating the previous choice across all subjects within each cohort. Figure 55 shows the relative frequencies of "staying" average across strong and weak roles for the two age groups. Young subjects' on average repeated his/her investment for about 44% of the time during entire experiments (42% as strong players and 46% as weak players), which is exactly the same probability of "staying" if a subject follows Nash equilibrium prediction perfectly throughout the experiment. Whereas the elderly subjects repeated his/her previous investment with significantly higher chance, around 60% of the time (55% as strong players and 66% as weak players).

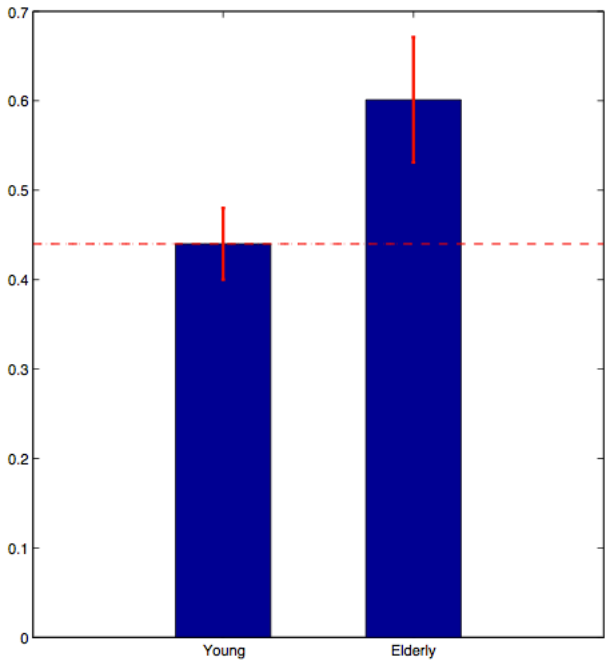

Figure 55: Comparison of probabilities of "staying" across different age groups. Red dash line is the probability of repeat the same investment is the subject follows Nash equilibrium prediction.

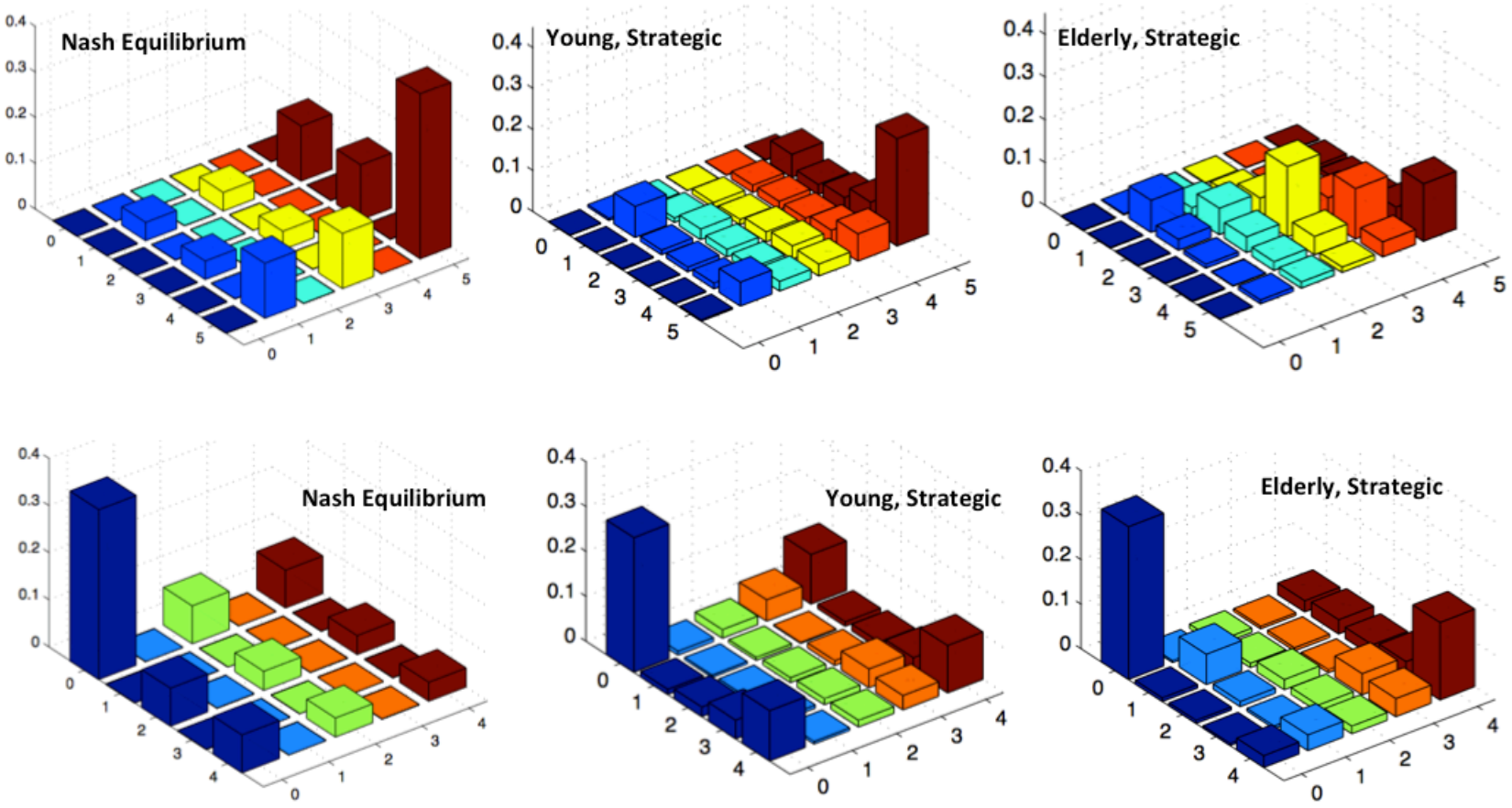

Figure 56: Transition matrix. On the X-axis of the matrix it is the empirical frequencies of the cohort's investment at trial t, and on the Y-axis, it is the empirical frequencies of the investment at trial t+1. UP: Strong players. Down: Weak Players.

Because "stickiness" is a very limited description for the overall dynamics of the play, next I explored the first order sequential dependencies of the choices by comparing transition matrices of the investment choice by each role at the group level. Figure 56 presents the joint empirical frequencies of investment choices at trial t+1 conditional investment at trial t for strong and weak roles. In both cases the transition matrices are computed by pooling the observations across trials and subjects within each age group. Again, as benchmarks for comparison, the transition matrices of a perfect Nash equilibrium player are included. The upper row in Figure 56 contains transition matrices for strong players. The young cohort randomized between investing 1 and 5 with higher chance of selecting 5 as predicted by Nash equilibrium. Yet they were discrepant from Nash equilibrium by mixing over investing 4 and 5 instead of including 3 into the randomization. The elderly cohort, on the other hand, has higher weight on the diagonal of the transition matrix, indicating they tended to repeat their previous choices a lot more often than mixing between different choices. Similar pattern was also found for the weak role. The young subject group switched between investing 0 and 4 about 10% of the time each, as predicted by Nash equilibrium, yet they underinvested 2 overinvested 3. The elderly subjects, however, mainly repeated investing 0, 1, 3 and 4, with occasional switching between 3 and 4. These results are consistent with the conjecture that the elderly subjects were on average further away from the mixed-strategy equilibrium solution that calls for perfect randomization of choices between successive trials.

Model-free measurement of strategic learning

In order to investigate into the reasons for the above discrepancies between younger and elder subjects behavior and Nash equilibrium predictions, I examined the level of strategic reasoning (theory of mind) within each cohort when playing the Patent Race game. The logic of iteratively elimination of dominated strategies is explained in the previous chapter.

Firstly, I compared the portions of subjects who have almost never played the iteratively dominated strategies throughout 80 rounds, which provides a necessary but not sufficient measure for the depth of strategic reasoning. Panels on the left column of Figure 57 present the empirical cumulative distributions for the portion of subjects playing the corresponding iteratively dominated strategies. For example, the left panel of Figure 57A suggests that almost all subjects in both young and elderly cohorts deleted investing 0 as strong player (86% in young cohort and 100% in old cohort), the strongly dominated strategy. Similarly, about half of the subjects in both age groups have never chosen to invest 1 as weak players (60% in young cohort and 56% in elderly cohort) as suggested by the left panel of Figure 57B. About 30% of subjects in both age groups have never played 0 and 2 as strong players (47% in young cohort and 38% in the elderly cohort) as shown in the left panel of Figure 57C. Finally, almost none of the subjects from both age groups have deleted both 1 and 3 as weak players.

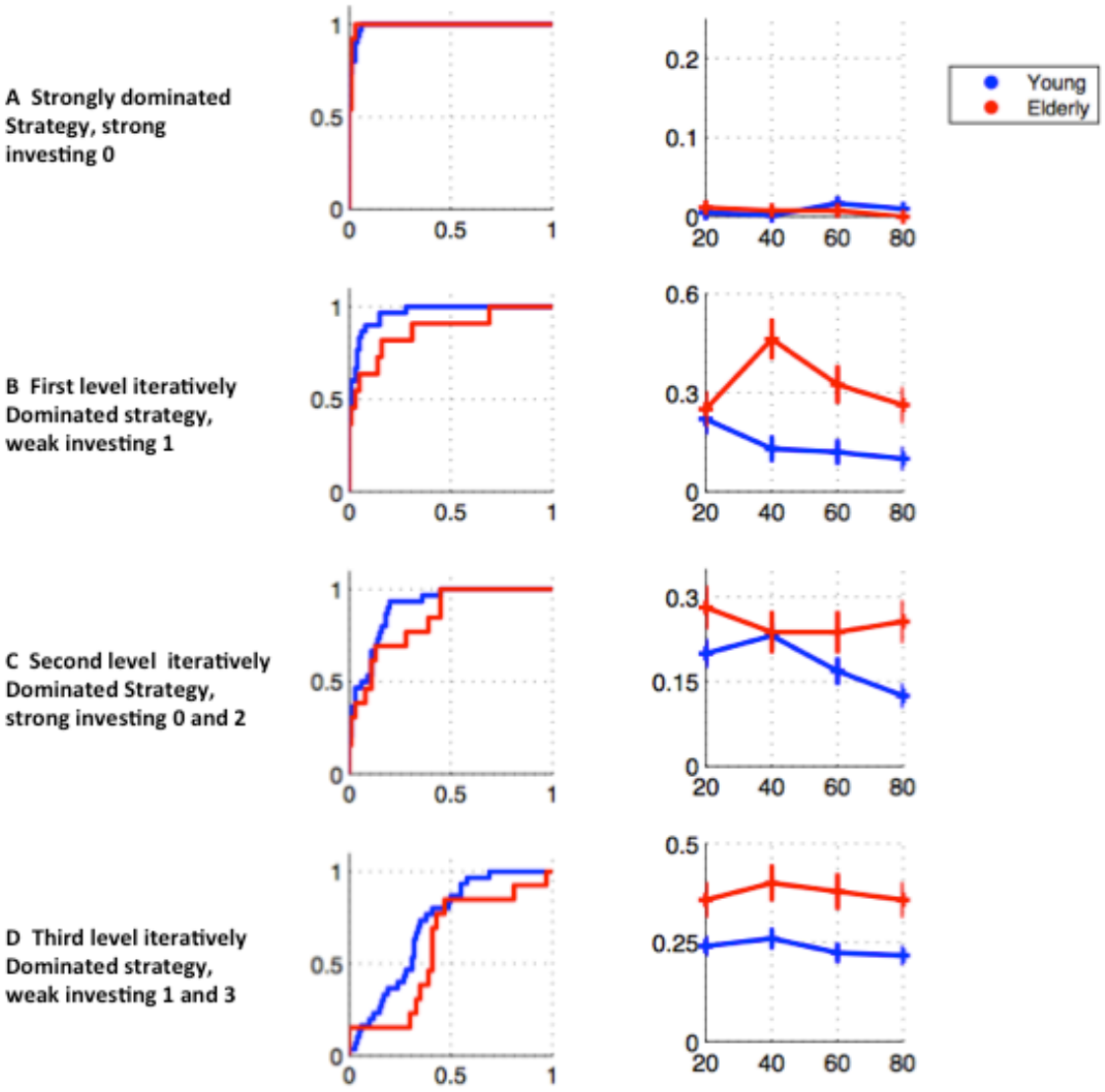

Figure 57: Iteratively elimination of strongly dominated strategies. Left: Empirical cumulative distributions of the portion of subjects playing the corresponding iteratively dominated strategies. X-axis indicates the frequency of playing the corresponding strategies. Y-axis represents the portion of experimental cohort. The starting point on the Y-axis hence represents the portion of the experimental cohort that has eliminated the corresponding strategy (strategies). Right: Time trend for playing the corresponding strategy (strategies) within the portion of subjects who failed to eliminate the strategy (strategies).

Taken together, these results indicate that (1) less and less subjects engaged in strategic reasoning as the depth of iterative dominance increased. (2) Consistent with the observations from many previous studies (Nagel, 1999), the level of reasoning for both age groups does not go beyond of level of 3; (3) no significant difference was found in the portion and depth of engaging in iterative reasoning between the young and elderly samples in our experiment, in line with the findings based on a one-shot "beauty contest" game for the elderly subjects (Figure 51, Kovalchik, 2005).

The above measurement takes a static view in interpretation subjects' choices, largely ignoring the underlying time trend and the possibility for subjects learning to engage in strategic reasoning. For example, it is possible that a subject does not initially realized investing 1 is an inferior option as a weak player, but gradually decreases the chance of selecting 1 as best responding to his/her observation that 0 has rarely been selected by the opponents as choices unfolded over time. In order to investigate the possible learning effect, I then evaluated the time trends in playing iteratively dominated strategies. In particular, I focused on the subset of subjects who failed to completely eliminate the iteratively (strongly) dominated strategies, divided 80 rounds of the game into 4 blocks each with 20 rounds, and then compared the frequencies of selecting those choices within each block for both young and elderly cohorts respectively. The results are shown in the right column of Figure 57. Interestingly, as indicated by Figure 57B, the elderly and young subjects who failed to eliminate investing 1 as weak player actually started with almost the same empirical frequency for selecting 1 during the first 20 rounds. In particular, both age groups started with a frequency close to the chance of playing pure randomly as weak players (20%). However, throughout 4 blocks, young

subjects significantly decreased the frequency of selecting 1, whereas the frequency of elderly cohort remained at similar level. Similar observation holds for strategies that are second level iteratively dominated, as suggested in the right panel of Figure 57C. Finally, both young and elderly subjects manifested no significant change in the frequencies of selecting both 1 and 3 as strong players (Figure 57D), the third level of iteratively dominated strategies.

Together these results suggested that although there exist no significant difference between young and elderly subjects' strategic reasoning from a static point of view, the two cohorts differ significantly in learning to perform strategically. According to our model-free measurement, elderly cohort presented a relatively flat learning curve. This finding is in line with previous studies based on IGT (Figure 52, Dohmen et al., 2005). However, such a model free measurement completely ignores the influence of opponents' choices, remains silent on the internal mechanism for how and why the elderly cohort has a flatter learning curve, and hence provides largely qualitative conclusions from the data. In order to quantify differences in the learning process, I introduced computational models aiming at capturing group differences parametrically in the next session.

4.6. Model-based Measurements

Computational Model

I adopted the "experience-weighted attraction" (EWA) model first introduced by Camerer and Ho (1999) to quantitatively compute the mapping from the stimulus inputs to the behavioral observations. There are two main attractive features of this model. First, it embeds two of the most widely used approaches to studying strategic learning in

competitive games – reinforcement learning and belief-based learning. The former is based on traditional models of reward learning (Schultz, Dayan et al. 1997; S. Sutton and G. Barto 1998; McClure, Berns et al. 2003). Players under this model are assumed to have a naïve mental model, and thus do *not* take into account of other players' behavior. Belief learning on the other hand, assumes more sophisticated players who use history of play to forecast other players' choices, and respond optimally given those forecasts (Cheung and Friedman 1997; Camerer, Ho et al. 2003), allowing them to learn to reduce playing the iteratively dominated strategies as suggested by the game theory concepts.

Second, it allows for a sophisticated way to account for the depreciation of the past by separating the discount of the past experience and the discount of the past subjective values. In this way, it is set up to capture subtle yet important distinguishes that may cause differences in learning rates. In particular, EWA assumes that two types of past experience are involved in strategic learning: the experience during the experiment (in-game experience) and the experience before participating the experiment (prior belief). Both two types of past experience are assumed to matter for all future choice decisions in different ways. In particular, EWA introduces separated decay rates that control the future strengths of influences by in-game experience and prior belief respectively. For example, for prior belief, EWA contains a free parameter (N(0)) for its initial strength, which decays over time at the rate of ρ, whereas in-game experience decays at a different rate of φ. Moreover, EWA allows for the nonlinear interactions among these parameters, providing a hybrid model that is superior to the linear combination of reinforcement and belief based learning.

Fro details of the EWA model and the meaning of the parameters, see chapter 2.

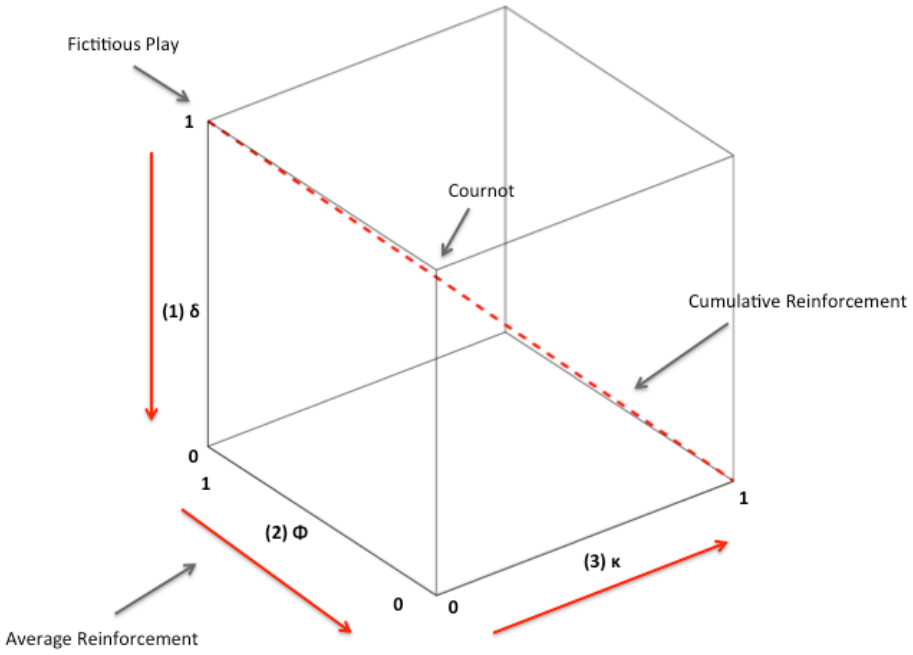

Figure 58: Hypotheses in the parameter space for EWA model. Elderly cohort is hypothesized to differ from the young cohort in the direction of one or more red arrows. Moreover, k is defined as $\kappa = 1 - (\frac{\rho}{\phi})$. Adapted from Camerer and Ho (1999).

Hypotheses

Based on observations from model-free measurements, I hypothesized that the difference in strategic learning can be captured by key parameters in my computational model across cohorts. In particular, one or more of the following may hold: (1) Old adults will employ less belief based learning strategies, and more reinforcement strategies. That is smaller δ for elderly cohort; (2) Old adults on average adapts slower because they discount their in-game experiences more rapidly, indicating they are insensitive to (easier to forget) the recent in-game experiences. That is a smaller φ for the elderly cohort; or (3) Elderly subjects are more "stubborn" in the sense that their pre-

game prior belief decays slower. That is the estimated value of $\kappa = 1 - (\frac{\rho}{\phi})$ for elderly cohort is closer to 0. Figure 59 visualizes the above three hypotheses.

Results

First I estimated group level parameters for both young and elderly cohorts based on EWA model. More specifically, I assumed that there exists a representative model with a single, shared set of parameters that can explain choices across all subjects' within the same age group possessing the same level of endowment. I thus pooled the observations conditional on age groups and roles, and fitted the choice data by maximizing the log likelihood of the observed choices over rounds for subject i. That is, $\sum_i \sum_t \log\left(p_i^{s_i(t)}(t)\right)$. To estimate the model, I conducted a grid search over a large range of values for all free parameters, since the likelihood function is not globally concave. Standard errors were estimated through standard jackknife procedure (Camerer and Ho 1999). Estimation results are summarized in Table 8 and visualized in Figure 59.

As hypothesized, elderly cohort has a relatively lower δ value, indicating that they are at group level they employ less belief based best responding strategy, and more simplistic reinforcement strategies. This is in line with the findings through fMRI that there may exist significant tissue loss in gray matter volume in mPFC (Figure 54), which is indicated involved in belief-based learning in my previous study.

		Group Level EWA Estimates				
Role	Group	δ	ρ	N_0	φ	λ
Strong	Young	.45	.90	1.48	.93	.46
	Elderly	.15	.62	2.54	.88	.23
Weak	Young	0.10	.95	8.24	.97	.58
	Elderly	0.00	.51	2.01	.89	.14

Table 8: EWA estimation results at cohort level. Parentheses contain standard error for the estimates obtained through standard jackknife procedure.

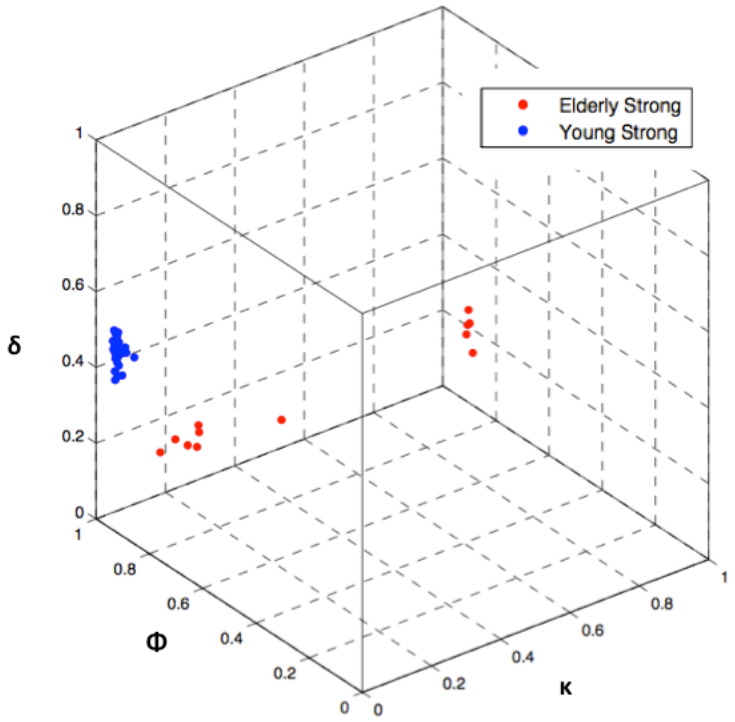

Figure 59: (A) Jackknife estimators for young and elderly cohorts playing as strong players respectively. (B) Jackknife estimators for young and elderly cohorts playing as weak players respectively.

In order to verify that the estimated different level of bias toward belief-based learning are indeed consistent with the model free measurements, we simply plot the estimated value of δ against empirical frequencies of playing the iteratively dominated strategies at individual subject level in Figure 60. It is clear from Figure 60 that for subjects who do play those iteratively dominated strategies with positive probabilities (i.e. the points that do not fall on the X-axis) , the individualized estimates of δ is negatively correlated with the empirical frequencies of selecting the 1^{st} and 2^{nd} order of iteratively dominated strategies, but not with the 3^{rd} level of interactively dominated strategies. It is consistent with the hypothesis that if a play adopt higher level of belief based learning, he will learn faster to eliminate at least low level of dominated strategies over 80 rounds of the play.

Furthermore, elderly cohort does not significantly differ from the young cohort in the in-game experience decay rate φ for both strong and weak roles. φ is usually associated with the speed of the adaption to the external environment. In general, if a player believes his/her opponent is a fast adaptor, he/she will adapt fast too.

That is, he/she will have a small φ that depreciates past attractions faster. According to our result both young and elderly cohorts utilized relatively large value of φ (.95 for young cohort and .89 for the elderly cohort average across strong and weak roles), suggesting that both groups played smoothly in response to the external environment depreciating the past in-game experience relatively slowly. Notably, our design controlled for the adaptation speed of the external environment by letting both young and elderly cohorts face the same pool of opponents' behavior.

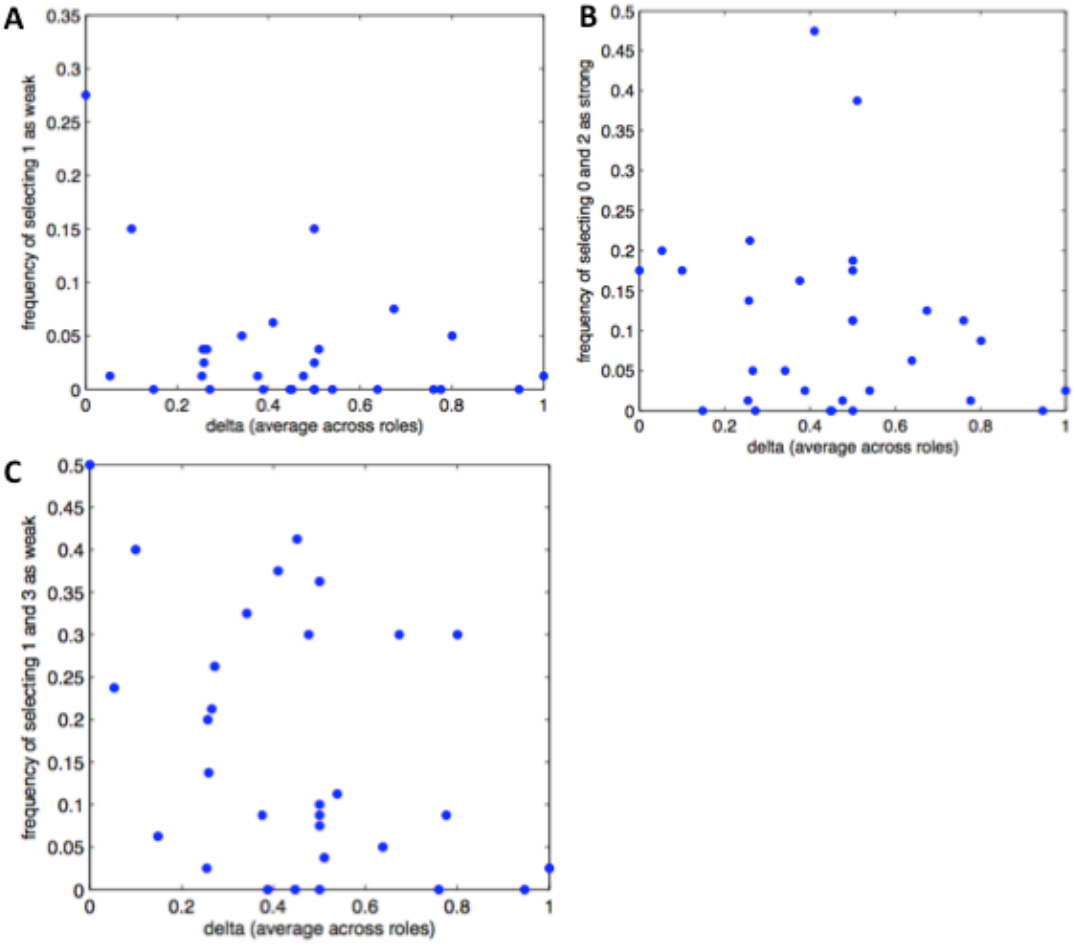

Figure 60: The relationship between estimated δ and frequency of playing the iteratively dominated strategies at individual subject level. A-C: estimated individualized δ against 1st, 2nd and 3rd level of elimination.

Interestingly, the estimated values for ρ are significantly different across two groups. ρ is the discount rate for the strength of experience measure N(t), and controls the influence of the prior beliefs. If ρ is large, prior belief will wear off quickly. (Camerer and Ho 1999) Our estimated result suggested that the elderly subjects have more persistent prior belief at group level. In order to visualize the result, Figure 61 presents simulated influence of prior belief over 80 rounds based on the above best fit estimation for each age group and each role. In all the cases, the influence of prior is measured by $\frac{N(0)}{N(t)}$, which is the weight assigned to A(0) within A(t). As benchmarks for comparisons, the time trends for $\frac{N(0)}{N(t)}$ under standard RL and standard fictitious play are also included. When RL is adopted, the prior will receive a time invariant weight that equals to 1. Whereas according to standard fictitious play, the influence of prior belief will drop to zero within the first 20 rounds. As shown in Figure 61, the elderly subjects' behavior is largely in line with reinforcement learning under the strong role, whereas young subjects' prior decays rapidly within first 20 rounds, similar to that in fictitious play. But its influence converges to .2 for the rest of the game. In weak role however, young subjects started with a much strong prior and converged to a similar level as that of the elderly subjects. Overall my result indicates that after controlling for N(0), the strength of the prior belief, prior belief has longer more persistent influence on elderly cohort's choices.

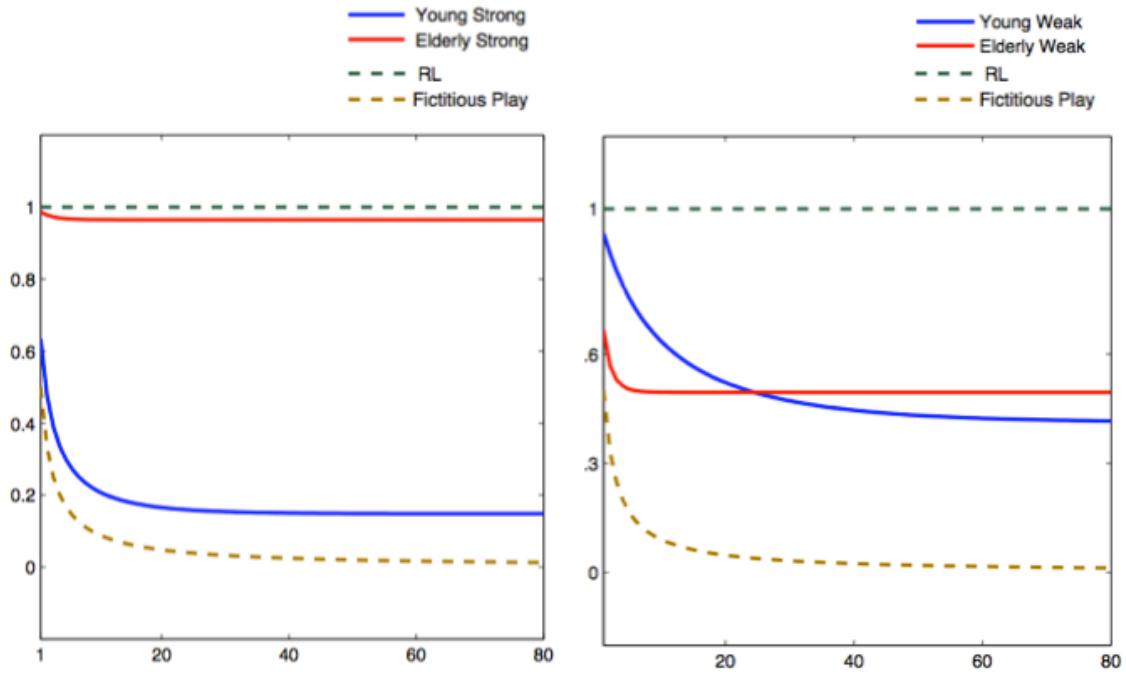

Figure 61: The simulated influences of prior belief throughout 80 rounds, measured by $\frac{N(0)}{N(t)}$, which is the weight assigned onto the initial attraction at time t. Left: In strong role. Right: In weak role.

To summarize, the computational model formalized the mechanical assumptions for internal reasoning and learning during repeated strategic interactions, seeking to identify the age-related difference for strategic learning, i.e. why at group level, the elderly subjects did not learn to eliminate the iteratively dominated strategies as the young cohort did. My result suggests that the reasons are two folds: on one hand, the elderly cohort adopts more reinforcement learning and less belief-based learning compared to the young cohort. Hence it is less likely for them to observe that a certain strategies have rarely been played by opponents, which in turn slows down learning to reduce the use of iteratively dominated strategies. On the other hand, elderly subjects on average have a more persistent prior belief, controlling for the strength of the prior.

Consequently their beliefs evolve slower over the course of the experiment, although they use the in-game experience as much as the young cohort.

4.7. Discussions

The study is motivated by my previous findings that strategic learning involves both striatal and medial prefrontal regions of the brain and the fact that elderly people usually experience degeneration in those brain regions. I used a repeated normal form Patent Race game as a probe to search for the possible behavioral signature underlying strategic learning for the elderly population. This is particularly interesting given the conflicting evidences in the existing aging literature.

In this study I compared the behavioral performance between the younger and older adults with both model free and model-based measurements. First I showed that consistent with previous studies, the elderly cohort and young cohort do not differ significantly in strategic reasoning (theory of mind) when assessing from a static point of view. However, from a dynamic point of view, the elderly subjects adapted poorer for the repeated strategic interactions compared to the young subjects. By adopting a computational framework that nests both reinforcement and belief based learning, and allows for sophisticated discount of past experience, I explored the internal mechanism for the age-related difference in strategic learning. Somewhat surprisingly, I showed that the elderly cohort on average learn slower, not because they forget the recent experiences faster than the young subjects as many stereotypes would suggest, but because (1)their choices evolve more through reinforcing the received rewards and punishments, while

less through best responding to the beliefs about the opponent's future actions; and (2) their prior belief decays slower and hence has more persistent influence in their decisions.

It is worth noting, however, that the study potentially suffers from the following limitations. Firstly, it lacks proper control over potential differences across age groups resulting from historical influences, such as educational opportunity, cultural factors and socioeconomic status. This kind of problem is almost unavoidable in almost all sectional studies on aging. Longitude studies, on the other hand, will be suitable for better control over historical factors.

Secondly, the group level EWA estimation is known to suffer from the downward biasing in estimated value for δ due to the possible individual heterogeneity in observed choices (Wilcox 2006). Such bias may magnify the estimated age-related difference in the engagement in reinforcement and belief-based learning, as the elderly cohort behavior may present larger individual differences. To alleviate the problem, individualized estimation is needed. Such individualized estimation will only serve as an initial step for developing biomarkers for the aging brain.

Future work will include the following. Firstly, I will compare the performances under computer vs. real human opponents treatments across young and elderly cohort with the same Patent Race game. My hypotheses are: (1) Even when computer algorithm adopts the exactly same probability as the empirical frequency of choices by the real human opponents, more naïve strategies will be used in response to the computer opponent. Elderly subjects on average will utilize lower level of strategic reasoning as measured by my model-free measurement and has smaller value of estimated δ as in the model-based measurement; (2) the difference across computer and real human opponents

treatments will be more significant within young cohort than the elderly cohort; (3) the elderly cohort learns slower in human treatment than in computer treatment, as their prior on how other human will behave have more persistent influences.

Secondly, I will characterize neural system underlying strategic learning and age-related changes, which will include using fMRI to examine the neural mechanisms underlying strategic learning in the two cohorts using the competitive game and will test the following two hypotheses derived from recent results in neuroeconomics. In particular, I have the following hypotheses: (1) strategic learning across all cohorts will be driven by multiple inputs from distinct brain systems—a striatum-based reinforcement learning system and a medial prefrontal cortex (mPFC) based belief-based learning system, consistent with my findings in Chpater 2. (2) Compared to young adults, mPFC activity associated with belief-based learning will be diminished in old adults relative to reinforcement learning in striatal regions. That is, the balance between reinforcement and social learning will tip toward reinforcement in old adults. Together, the proposed future work will aim at providing new measures for neuropsychological assessment of cognitive function and development of better tools for assessing cognitive function; improving our understanding of nervous system and behavioral changes that occur with normal aging and how brain function is maintained and enhanced; and identifying neuroimaging biomarkers for early detection of cognitive decline.

References

Amodio, D. M. and C. D. Frith (2006). "Meeting of minds: the medial frontal cortex and social cognition." Nature Reviews Neuroscience **7**(4): 268-277.

Badre, D. and A. Wagner (2004). "Selection, Integration, and Conflict Monitoring:: Assessing the Nature and Generality of Prefrontal Cognitive Control Mechanisms." Neuron **41**(3): 473-487.

Baier, B., H.-O. Karnath, et al. (2010). "Keeping memory clear and stable--the contribution of human basal ganglia and prefrontal cortex to working memory." Journal of neuroscience **30**(29): 9788-9792.

Barraclough, D., M. Conroy, et al. (2004). "Prefrontal cortex and decision making in a mixed-strategy game." Nat Neurosci **7**(4): 404-410.

Bechara, A. (2000). "Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions." Brain **123**(11): 2189-2202.

Bechara, A., D. Tranel, et al. (1997). "Deciding Advantageously Before Knowing the Advantageous Strategy." Science **275**(5304): 1293-1295.

Behrens, T., L. Hunt, et al. (2009). "The computation of social behavior." Science.

Behrens, T., L. Hunt, et al. (2008). "Associative learning of social value." NATURE **456**(7219): 245.

Bellebaum, C., B. Koch, et al. (2008). "Focal basal ganglia lesions are associated with impairments in reward-based reversal learning." Brain **131**(3): 829-841.

Bernheim, B. (2009). "On the potential of neuroeconomics: a critical (but hopeful) appraisal." American Economic Journal: Microeconomics **1**(2): 1-41.

Berridge, K. (1998). "What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience?" Brain Research Reviews.

Bhatt, M. A., T. Lohrenz, et al. (2010). "Neural signatures of strategic types in a two-person bargaining game." Proceedings of the National Academy of Sciences **107**(46): 19720.

Buckley, M., F. Mansouri, et al. (2009). "Dissociable Components of Rule-Guided Behavior Depend on Distinct Medial and Prefrontal Regions." Science **325**(5936): 52-58.

Bucks, B. K., A. B. Kennickell, et al. (2009). "Changes in US family finances from 2004 to 2007: Evidence from the Survey of Consumer Finances." Federal Reserve Bulletin **95**(1): A1-A55.

Camerer, C. F. (2008). "Neuroeconomics: Opening the Gray Box." Neuron **60**(3): 416-419.

Camerer, C. F. (2010). "Behavioral game theory: experiments in strategic interaction." BOOK: 1-248.

Camerer, C. F. and T. Ho (1999). "Experienced-weighted attraction learning in normal form games." Econometrica **67**(4): 827-874.

Camerer, C. F., T. Ho, et al. (2004). "A cognitive hierarchy model of games." Quarterly journal of Economics **119**(3): 861-898.

Camerer, C. F., T. Ho, et al. (2003). "Models of Thinking, Learning, and Teaching in Games." The American Economic Review **93**(2): 192-195.

Camille, N., G. Coricelli, et al. (2004). "The involvement of the orbitofrontal cortex in the experience of regret." Science.

Chamberlain, S., L. Menzies, et al. (2008). "Orbitofrontal dysfunction in patients with obsessive-compulsive disorder and their unaffected relatives." Science **321**(5887): 421-422.

Charlton, R. A., T. R. Barrick, et al. (2009). "Theory of mind associations with other cognitive functions and brain imaging in normal aging." Psychology and Aging **24**(2): 338.

Cheung, Y. and D. Friedman (1997). "Individual learning in normal form games: Some laboratory results." Games and Economic Behavior.

Cheung, Y. and D. Friedman (1998). "A comparison of learning and replicator dynamics using experimental data." Journal of Economic Behavior and ….

Cheung, Y.-W. and D. Friedman (1997). "Individual Learning in Normal Form Games: Some Laboratory Results." Games and Economic Behavior **19**: 46-76.

Coricelli, G. and R. Nagel (2009). "Neural correlates of depth of strategic reasoning in medial prefrontal cortex." Proceedings of the ….

Crawford, G. and M. Shum (2005). "Uncertainty and learning in pharmaceutical demand." Econometrica.

Daw, N. D., J. P. O'Doherty, et al. (2006). "Cortical substrates for exploratory decisions in humans." NATURE.

Delgado, M., A. Schotter, et al. (2008). "Understanding Overbidding: Using the Neural Circuitry of Reward to Design Economic Auctions." Science **321**(5897): 1849.

Denburg, N., D. Tranel, et al. (2005). "The ability to decide advantageously declines prematurely in some normal older persons." Neuropsychologia **43**(7): 1099-1106.

Denburg, N. L., C. A. Cole, et al. (2007). "The Orbitofrontal Cortex, Real‐World Decision Making, and Normal Aging." Annals of the New York Academy of Sciences **1121**(1): 480-498.

Dennis, N. and R. Cabeza (2008). "Neuroimaging of healthy cognitive aging." Handbook of aging and cognition.

Dorris, M. and P. W. Glimcher (2004). "Activity in Posterior Parietal Cortex Is Correlated with the Relative Subjective Desirability of Action." Neuron **44**(2): 365-378.

Duffy, J. and R. Nagel (1997). "On the Robustness of Behaviour in Experimental `Beauty Contest' Games." Economic Journal **107**(445): 1684-1700.

Erev, I. and A. Rapoport (1998). "Coordination,"Magic," and Reinforcement Learning in a Market Entry Game." Games and Economic Behavior **23**(2): 146-175.

Erev, I. and A. Roth (1998). "Predicting how people play games: Reinforcement learning in experimental …." The American Economic Review.

Erev, I., A. Roth, et al. (2007). "Learning and equilibrium as useful approximations: Accuracy of prediction on …." Economic Theory.

Erixonlindroth, N., L. Farde, et al. (2005). "The role of the striatal dopamine transporter in cognitive aging." Psychiatry Research: Neuroimaging **138**(1): 1-12.

Fehr, E. and C. F. Camerer (2007). "Social neuroeconomics: the neural circuitry of social preferences." TRENDS in Cognitive Sciences: 1-9.

Feltovich, N. (2000). "Reinforcement-based vs. belief-based learning models in experimental asymmetric- …." Econometrica.

Fiorillo, C., P. Tobler, et al. (2003). "Discrete coding of reward probability and uncertainty by dopamine neurons." Science **299**(5614): 1898-1902.

Foster, D. and R. Vohra (1998). "Asymptotic Calibration." Biometrika **85**(2): 379-390.

Foster, D. and H. P. Young (2006). "Regret testing: learning to play Nash equilibrium without knowing you have an opponent." Theoretical Economics: 1-27.

Frank, M. J. (2005). "Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism." Journal of Cognitive Neuroscience **17**(1): 51-72.

Frank, M. J., J. Samanta, et al. (2007). "Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism." Science **318**(5854): 1309-1312.

Frank, M. J., L. Seeberger, et al. (2004). "By carrot or by stick: cognitive reinforcement learning in parkinsonism." Science **306**(5703): 1940-1943.

Frank, M. J., B. Woroch, et al. (2005). "Error-related negativity predicts reinforcement learning and conflict biases." Neuron **47**(4): 495-501.

Fudenberg, D. and D. K. K. Levine (1998). "The theory of learning in games." BOOK: 276.

Fudenberg, D. and D. K. K. Levine (2009). "Learning and Equilibrium." Annu. Rev. Econ. **1**(1): 385-420.

Fudenberg, D. and D. Kreps (1993). "Learning mixed equilibria." Games and Economic Behavior.

Grady, C. (2008). "Cognitive neuroscience of aging." Ann NY Acad Sci **1124**: 127-144.

Green, L., A. Fry, et al. (1994). "Discounting of Delayed Rewards: A Life-Span Comparison." Psychol Sci **5**(1): 33-36.

Grossmann, I., J. Na, et al. (2010). "Reasoning about social conflicts improves into old age." PNAS **107**(16): 7246-7250.

Gul, F. and W. Pesendorfer (2008). "The case for mindless economics." The foundations of Positive and normative Economics: A handbook: 3–42.

Gul, F. and W. Pesendorfer (2009). "A Comment on Bernheim's Appraisal of Neuroeconomics." American Economic Journal: Microeconomics **1**(2): 42-47.

Hampton, A., P. Bossaerts, et al. (2008). "Neural correlates of mentalizing-related computations during strategic interactions in humans." Proceedings of the … **105**(18): 6741.

Happé, F., E. Winner, et al. (1998). "The getting of wisdom: theory of mind in old age." Dev Psychol **34**(2): 358-362.

Hare, T., C. F. Camerer, et al. (2009). "Self-Control in Decision-Making Involves Modulation of the vmPFC Valuation System." Science **324**(5927): 646-648.

Harrison, G. (2008). "Neuroeconomics: A Critical Reconsideration." Economics and Philosophy **24**(03): 303-344.

Hart, S. and A. Mas-Colell (2003). "Regret-based continuous-time dynamics." Games and Economic Behavior.

Haruno, M. and M. Kawato (2009). "Activity in the Superior Temporal Sulcus Highlights Learning Competence in an Interaction Game." Journal of neuroscience **29**(14): 4542-4547.

Hayden, B., J. Pearson, et al. (2009). "Fictive reward signals in the anterior cingulate cortex." Science.

Hedden, T. and J. D. E. Gabrieli (2004). "Insights into the ageing mind: a view from cognitive neuroscience." Nature Reviews Neuroscience **5**(2): 87-96.

Heinemann, F., R. Nagel, et al. (2009). "Measuring strategic uncertainty in coordination games." Review of Economic Studies **76**(1): 181-221.

Ho, T., C. F. Camerer, et al. (2007). "Self-tuning experience weighted attraction learning in games." Journal of Economic Theory.

Hofbauer, J. and K. Sigmund (1998). "Evolutionary games and population dynamics." BOOK: 323.

Hsu, M., M. Bhatt, et al. (2005). "Neural systems responding to degrees of uncertainty in human decision-making." Science.

Huck, S., H.-T. Normann, et al. (1999). "Learning in Cournot Oligopoly - An Experiment." Economic Journal **109**(454): C80-C95.

Jenkins, A. C. and J. P. Mitchell (2010). "Mentalizing under uncertainty: dissociated neural responses to ambiguous and unambiguous mental state inferences." Cereb Cortex **20**(2): 404-410.

Kable, J. and P. W. Glimcher (2007). "The neural correlates of subjective value during intertemporal choice." Nature neuroscience **10**(12): 1625-1633.

Knoch, D., A. Pascual-Leone, et al. (2006). "Diminishing Reciprocal Fairness by Disrupting the Right Prefrontal Cortex." Science **314**(5800): 829-832.

Knutson, B., R. Momenan, et al. (2001). "Negative association of neuroticism with brain volume ratio in healthy humans." Biol Psychol **50**(9): 685-690.

Koenigs, M., L. Young, et al. (2007). "Damage to the prefrontal cortex increases utilitarian moral judgements." NATURE **446**: 908-911.

Kovalchik, S., C. F. Camerer, et al. (2005). "Aging and decision making: A comparison between neurologically healthy elderly and young individuals." Journal of Economic Behavior & Organization **58**(1): 79-94.

Krajbich, I., R. Adolphs, et al. (2009). "Economic Games Quantify Diminished Sense of Guilt in Patients with Damage to the Prefrontal Cortex." J Neurosci **29**(7): 2188-2192.

Krueger, F., J. Grafman, et al. (2008). "Neural correlates of economic game playing." Philos Trans R Soc Lond, B, Biol Sci **363**(1511): 3859-3874.

Lee, D. (2008). "Game theory and neural basis of social decision making." Nature neuroscience.

Lee, D., B. McGreevy, et al. (2005). "Learning and decision making in monkeys during a rock-paper-scissors game." Brain Res Cogn Brain Res **25**(2): 416-430.

Li, S.-C., G. Biele, et al. (2007). "Aging and neuroeconomics: insights from research on neuromodulation of reward-based decision making." Analyse & Kritik **29**: 97-111.

Lohrenz, T., K. McCabe, et al. (2007). "Neural signature of fictive learning signals in a sequential investment task." Proceedings of the … **104**(22): 9493.

MacDonald, A., J. Cohen, et al. (2000). "Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control." Science **288**(5472): 1835.

Maia, T. and J. McClelland (2004). "A reexamination of the evidence for the somatic marker hypothesis: What participants really know in the Iowa gambling task." Proceedings of the … **101**(45): 16075.

Maia, T. V. and M. J. Frank (2011). "From reinforcement learning models to psychiatric and neurological disorders." Nature Publishing Group **14**(2): 154-162.

Mansouri, F. A., M. J. Buckley, et al. (2007). "Mnemonic Function of the Dorsolateral Prefrontal Cortex in Conflict-Induced Behavioral Adjustment." Science **318**(5852): 987-990.

Marschner, A., T. Mell, et al. (2005). "Reward-based decision-making and aging." Brain Research Bulletin **67**(5): 382-390.

McClure, S., G. Berns, et al. (2003). "Temporal prediction errors in a passive learning task activate human striatum." Neuron.

McClure, S., K. Ericson, et al. (2007). "Time discounting for primary rewards." J Neurosci **27**(21): 5796-5804.

McClure, S., D. Laibson, et al. (2004). "Separate neural systems value immediate and delayed monetary rewards." Science **306**(5695): 503-507.

McKelvey, R. and T. Palfrey (1992). "An Experimental Study of the Centipede Game." Econometrica **60**(4): 803-836.

McKinnon, M. C. and M. Moscovitch (2007). "Domain-general contributions to social reasoning: theory of mind and deontic reasoning re-explored." Cognition **102**(2): 179-218.

Mell, T., H. Heekeren, et al. (2005). "Effect of aging on stimulus-reward association learning." Neuropsychologia **43**(4): 554-563.

Miller, E. (2000). "The prefontral cortex and cognitive control." Nature Reviews Neuroscience **1**(1): 59-65.

Milner, B. (1963). "Effects of different brain lesions on card sorting: The role of the frontal lobes." Archives of Neurology.

Mookherjee, D. and B. Sopher (1994). "Learning behavior in an experimental matching pennies game." Games and Economic Behavior **7**(1): 62-91.

Mookherjee, D. and B. Sopher (1997). "Learning and decision costs in experimental constant sum games." Games and Economic Behavior.

Nagel, R. (1995). "Unraveling in Guessing Games: An Experimental Study." The American Economic Review **85**(5): 1313-1326.

Narayanan, S. and P. Manchanda (2009). "Heterogeneous learning and the targeting of marketing communication for new products." Marketing Science **28**(3): 424-441.

Nyberg, L., J. Sandblom, et al. (2003). "Neural correlates of training-related memory improvement in adulthood and aging." PNAS **100**(23): 13728-13733.

Nyhus, E. and F. Barceló (2009). "The Wisconsin Card Sorting Test and the cognitive assessment of prefrontal executive functions: A critical update." Brain and Cognition **71**(3): 437-451.

O'Doherty, J. P., T. Buchanan, et al. (2006). "Predictive neural coding of reward preference involves dissociable responses in human ventral midbrain and ventral striatum." Neuron **49**(1): 157-166.

O'Doherty, J. P., H. Critchley, et al. (2003). "Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices." J Neurosci **23**(21): 7931-7939.

O'Doherty, J. P., P. Dayan, et al. (2003). "Temporal difference models and reward-related learning in the human brain." Neuron.

O'Doherty, J. P., P. Dayan, et al. (2004). "Dissociable roles of ventral and dorsal striatum in instrumental conditioning." Science **304**(5669): 452-454.

Olds, J. and P. Milner (1954). "Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain." J Comp Physiol Psychol **47**(6): 419-427.

Ongür, D. and J. L. Price (2000). "The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans." Cereb Cortex **10**(3): 206-219.

Padoa-Schioppa, C. and J. Assad (2006). "Neurons in the orbitofrontal cortex encode economic value." NATURE **441**(7090): 223-226.

Rangel, A., C. F. Camerer, et al. (2008). "A framework for studying the neurobiology of value-based decision making." Nature Reviews Neuroscience.

Rapoport, A. and W. Amaldoss (2000). "Mixed strategies and iterative elimination of strongly dominated strategies: An experimental investigation of states of knowledge." Journal of Economic Behavior & Organization **42**(4): 483-521.

Rappaport, A. M. and M. Dragut (2005). "High-Age Implications of Postretirement Risks." Benefits quarterly **21**(4): 12.

Resnick, S. M., D. L. Pham, et al. (2003). "Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain." J Neurosci **23**(8): 3295-3301.

Reuter-Lorenz, P. and C. Lustig (2005). "Brain aging: reorganizing discoveries about the aging mind." Current opinion in neurobiology **15**(2): 245-251.

Rorden, C. and M. Brett (2000). "Stereotaxic display of brain lesions." Behavioural neurology **12**(4): 191-200.

Roth, A. and I. Erev (1995). "… in extensive-form games: Experimental data and simple dynamic models in the intermediate …." Games and Economic Behavior.

Roth, A. and I. Erev (1995). "Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term*." Games and Economic Behavior **8**(1): 164-212.

Rudebeck, P. H. and E. A. Murray (2008). "Amygdala and Orbitofrontal Cortex Lesions Differentially Influence Choices during Object Reversal Learning." Journal of neuroscience **28**(33): 8338-8343.

Rustichini, A. (2009). "Is There a Method of Neuroeconomics?" American Economic Journal: Microeconomics **1**(2): 48-59.

S. Sutton, R. and A. G. Barto (1998). "Reinforcement learning: an introduction." BOOK: 322.

Salmon, T. (2001). "An evaluation of econometric models of adaptive learning." Econometrica.

Samanez-Larkin, G., C. Kuhnen, et al. (2010). "Variability in nucleus accumbens activity mediates age-related suboptimal financial risk taking." J Neurosci **30**(4): 1426-1434.

Samanez-Larkin, G. R., S. E. B. Gibbs, et al. (2007). "Anticipation of monetary gain but not loss in healthy older adults." Nat Neurosci **10**(6): 787-791.

Sanfey, A. G., G. Loewenstein, et al. (2006). "Neuroeconomics: cross-currents in research on decision-making." Trends in Cognitive Sciences **10**(3): 108-116.

Sanfey, A. G., J. Rilling, et al. (2003). "The neural basis of economic decision-making in the ultimatum game." Science.

Sargent, T. C. a. T. J. (2006). "The conquest of U.S. inflation: learning and robustness to model uncertainty." 1-57.

Saxe, R. and J. Haushofer (2008). "For Love or Money: A Common Neural Currency for Social and Monetary Reward." Neuron **58**(2): 164-165.

Schaie, K. W. (1996). Intellectual development in adulthood: The Seattle longitudinal study.

Schott, B. H., L. Niehaus, et al. (2007). "Ageing and early-stage Parkinson's disease affect separable neural mechanisms of mesolimbic reward processing." Brain **130**(9): 2412-2424.

Schultz, W. (2000). "Multiple reward signals in the brain." Nat Rev Neurosci **1**(3): 199-207.

Schultz, W., P. Dayan, et al. (1997). "A neural substrate of prediction and reward." Science **275**(5306): 1593-1599.

Slessor, G., L. H. Phillips, et al. (2007). "Exploring the specificity of age-related differences in theory of mind tasks." Psychology and Aging **22**(3): 639-643.

Sobel, J. (2009). "Neuroeconomics: A Comment on Bernheim." American Economic Journal: Microeconomics **1**(2): 60-67.

Stahl, D. O. (1996). "Boundedly rational rule learning in a guessing game." Games and Economic Behavior **16**: 303-330.

Staudinger, U. M., S. W. Cornelius, et al. (1989). "The aging of intelligence: Potential and limits." The ANNALS of the American Academy of Political and Social Science **503**(1): 43.

Stone, V. E., S. Baron-Cohen, et al. (1998). "Frontal lobe contributions to theory of mind." J Cogn Neurosci **10**(5): 640-656.

Tekin, S. and J. L. Cummings (2002). "Frontal-subcortical neuronal circuits and clinical neuropsychiatry: an update." J Psychosom Res **53**(2): 647-654.

Templeton, V. H. and D. N. Kirkman (2007). "Fraud, vulnerability, and aging: Case studies." Alzheimer's Care Today **8**(3): 265.

Thevarajah, D., A. Mikulic, et al. (2009). "Role of the Superior Colliculus in Choosing Mixed-Strategy Saccades." J Neurosci **29**(7): 1998-2008.

United Nation, D. o. and S. A. P. Economic Division (2001). "WORLD POPULATION AGEING: 1950-2050."

W. Glimcher, P., C. F. Camerer, et al. (2008). "Neuroeconomics: decision making and the brain." BOOK: 538.

Walhovd, K., A. Fjell, et al. (2005). "Effects of age on volumes of cortex, white matter and subcortical structures." Neurobiology of Aging **26**(9): 1261-1270.

Wallis, J. D. (2007). "Orbitofrontal Cortex and Its Contribution to Decision-Making." Annu. Rev. Neurosci. **30**(1): 31-56.

Wilcox, N. (2006). "Theories of learning in games and heterogeneity bias." Econometrica: 1271-1292.

Williams, L., K. Brown, et al. (2006). "The Mellow Years?: Neural Basis of Improving Emotional Stability over Age." J Neurosci **26**(24): 6422-6430.

Yamada, H., N. Matsumoto, et al. (2004). "Tonically Active Neurons in the Primate Caudate Nucleus and Putamen Differentially Encode Instructed Motivational Outcomes of Action." J Neurosci **24**(14): 3500.

Yehene, E., N. Meiran, et al. (2008). "Basal ganglia play a unique role in task switching within the frontal-subcortical circuits: evidence from patients with focal lesions." J Cogn Neurosci **20**(6): 1079-1093.

Yoshida, W., I. Dziobek, et al. (2010). "Cooperation and heterogeneity of the autistic mind." The Journal of ….

Young, L., F. Cushman, et al. (2007). "The neural basis of the interaction between theory of mind and moral judgment." PNAS **104**(20): 8235-8240.